

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 04:26:11 ; Search time 59 seconds
(without alignments)
3701.657 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKIIFLCSEFLFIINTQCV.....SNFLGISFLILMLLYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

A_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8418	99.9	1639	19 AAW54145	P. falciparum synt
2	8139.5	96.6	1654	6 AAP50777	Sequence of the P1
3	1881	22.3	355	20 AAY09372	Merozoite surface
4	1881	22.3	355	20 AAY05832	Merozoite surface
5	1881	22.3	361	20 AAY09373	Merozoite surface
6	1881	22.3	361	20 AAY05833	Merozoite surface
7	1869	22.2	376	20 AAY09374	Modified merozoite
8	1869	22.2	376	20 AAY05834	Modified merozoite
9	1333	15.8	262	22 AAB37610	Merozoite surface
10	1165.5	13.8	375	22 AAB83926	A major merozoite

11	913	10.8	355	9 AAP80547	Polypeptide p190-3
12	910.5	10.8	219	9 AAP82523	31-1 Repeated Dele
13	905	10.7	219	9 AAP80544	Polypeptide p190-1
14	794	9.4	244	6 AAP50304	Storage-specific,
15	677	8.0	151	9 AAP80546	Polypeptide p190-2
16	637	7.6	116	18 AAW36103	PfMSPI(p19)A prote
17	637	7.6	116	18 AAW22527	PfMSPI(p19)A prote
18	584	6.9	115	23 AAM50527	Plasmodium falci
19	560	6.6	108	22 AAB37609	Merozoite surface
20	559.5	6.6	1979	21 AAB18171	Plasmodium falci
21	556	6.6	96	22 AAB37608	Merozoite surface
22	539	6.4	127	18 AAW22593	PfMSPI(p19)S prote
23	539	6.4	127	18 AAW36102	PfMSPI(p19)S prote
24	455.5	5.4	1254	11 AAR07503	Merozoite apical-en
25	455.5	5.4	1254	18 AAW24575	Merozoite apical-e
26	453.5	5.4	2485	21 AAB18172	Plasmodium falci
27	450	5.3	5024	22 AAG82935	S. epidermidis ope
28	448	5.3	10182	23 ABF38314	Staphylococcus epi
29	445	5.3	2633	22 ABG06505	Novel human diago
30	444.5	5.3	2688	22 AAW40883	Human polypeptide
31	442.5	5.3	3973	21 AAB18253	Human polypeptide
32	441	5.2	2663	22 AAM39097	Plasmodium falci
33	431	5.1	980	21 AAB18294	Plasmodium falci
34	425.5	5.1	2037	22 AAG06301	Novel human diago
35	424.5	5.0	206	22 AAU00669	Plasmodium vivax m
36	424	5.0	102	16 AAP50303	Storage-specific,
37	414	4.9	2010	21 AAB18218	Plasmodium falci
38	411.5	4.9	1516	21 AAB18195	Plasmodium falci
39	408.5	4.8	2539	21 AAB18198	Plasmodium falci
40	402.5	4.8	2184	22 AAE00425	P. falciparum telo
41	399	4.7	2013	21 AAB18265	Plasmodium falci
42	393	4.7	2954	20 AAY01632	Amino acid sequenc
43	388.5	4.6	2295	21 AAB18180	Plasmodium falci
44	387.5	4.6	4134	20 AAY31946	Plasmodium falci
45	387	4.6	3696	23 ABP40235	Staphylococcus epi
46	383.5	4.6	1308	21 AAB18167	Plasmodium falci
47	383.5	4.6	2573	21 AAB18234	Plasmodium falci
48	379.5	4.5	2206	21 AAB18254	Plasmodium falci
49	375	4.5	1780	22 AAB38681	Human polypeptide
50	372	4.4	2500	21 AAB18272	Plasmodium falci
51	370	4.4	1411	17 AAW02258	Nucleolar/endosoma
52	369.5	4.4	1807	22 AAB85697	Recombinant protei
53	369	4.4	1087	20 AAY19935	B. burgdorferi ant
54	369	4.4	1119	20 AAY19934	B. burgdorferi ant
55	368.5	4.4	1788	22 AAW40467	Human polypeptide
56	365.5	4.3	2380	21 AAB18315	Plasmodium falci
57	363	4.3	2346	22 AAB63519	Drosophila melanog
58	362.5	4.3	1712	21 AAB18205	Plasmodium falci
59	360.5	4.3	1714	21 AAB18275	Plasmodium falci
60	359	4.3	2415	22 AAG20279	Novel human diago
61	357.5	4.2	1817	21 AAB18255	Plasmodium falci
62	356.5	4.2	2400	22 AAG20278	Novel human diago
63	356.5	4.2	3248	17 AAR99795	Kinetochore protei
64	354	4.2	1690	22 ABB61144	Drosophila melanog
65	354	4.2	1690	22 ABB61173	Drosophila melanog
66	354	4.2	1997	21 AAB18287	Plasmodium falci
67	353.5	4.2	2285	18 AAW98149	Bacillus subtilis
68	351	4.2	1786	18 AAW24790	P. falciparum live
69	351	4.2	1787	23 AAB96699	Plasmodium falci
70	350.5	4.2	1177	22 AAB96721	Putative P. abysssi
71	350.5	4.2	2482	16 AAR72826	Human mitosisin. Ho
72	350.5	4.2	2482	19 AAW23996	Human mitosisin amin
73	350	4.2	1802	21 AAB18217	Plasmodium falci
74	345	4.1	1817	21 AAB18301	Plasmodium falci
75	343	4.1	1939	23 ABB77096	Human alpha-myosin
76	342.5	4.1	1489	22 ABB59948	Drosophila melanog
77	341	4.0	2028	22 AAB85698	Recombinant protei
78	340	4.0	2816	22 AAB68572	Human novel cytoki
79	338	4.0	1886	19 AAW54241	Rattus norvegicus
80	336.5	4.0	1743	19 AAW98879	H. pylori GHPO 175
81	336.5	4.0	2013	22 ABB62322	Drosophila melanog
82	336.5	4.0	2067	22 ABB71125	Drosophila melanog
83	335	4.0	3685	10 AAP90373	Sequence encoded b

84	334.5	4.0	2918	22	ABG27218	Novel human diagno
85	334	4.0	1392	20	AAV06999	Restin protein seq
86	333	4.0	1931	22	ABB61012	Drosophila melanog
87	332	3.9	976	22	AAG66581	Human SCP-1 mutein
88	330	3.9	3685	10	AAP90290	Human Duchenne mus
89	330	3.9	7201	22	ABB71136	Drosophila melanog
90	329.5	3.9	1358	21	AAB18324	Plasmodium falcipa
91	329.5	3.9	2056	22	ABB59344	Drosophila melanog
92	329	3.9	996	22	AAU33755	Staphylococcus aur
93	329	3.9	1009	22	AAU36548	Plasmodium falcipa
94	329	3.9	1346	21	AAB18236	Plasmodium falcipa
95	329	3.9	2441	21	AAB18161	Plasmodium falcipa
96	328	3.9	2496	21	AAB18222	Plasmodium falcipa
97	327.5	3.9	2172	22	ABB67958	Drosophila melanog
98	324.5	3.9	2096	22	AAB41592	Human ORFX ORF1356
99	323.5	3.8	2139	22	ABA47278	PN7711 Homo sapi
100	323.5	3.8	2143	22	ABG01716	Novel human diagno
101	322.5	3.8	1427	12	AAU10534	Human 160kD mediat
102	320	3.8	1095	22	AGB83030	S. epidermidis ope
103	320	3.8	1988	22	RAM40999	Human polypeptide
104	320	3.8	1988	22	RAM40999	Human polypeptide
105	320	3.8	2779	22	ABG62371	Drosophila melanog
106	319.5	3.8	1279	22	AGB83047	S. epidermidis ope
107	319.5	3.8	2659	23	AAU75489	S. aureus antigeni
108	319	3.8	1436	21	AAU18199	Plasmodium falcipa
109	318	3.8	1090	23	ABB54799	Lactococcus lactis
110	317.5	3.8	1121	21	AAB18241	Plasmodium falcipa
111	316.5	3.8	3158	22	AAU37018	Staphylococcus aur
112	316	3.8	6815	22	ABB66811	Drosophila melanog
113	315.5	3.7	1948	22	ABG21233	Novel human diagno
114	314.5	3.7	1151	23	ABP38397	Staphylococcus epi
115	312.5	3.7	1833	22	ABB71141	Drosophila melanog
116	312.5	3.7	2048	22	AAU40027	Human polypeptide
117	312	3.7	1409	22	ABB65706	Drosophila melanog
118	312	3.7	2835	23	AAU75100	Novel signal pathw
119	310	3.7	1211	23	ABP39975	Staphylococcus epi
120	310	3.7	1526	22	ABG07893	Novel human diagno
121	308.5	3.7	2059	23	ABP25711	Streptococcus poly
122	308	3.7	696	21	AAB18181	Plasmodium falcipa
123	308	3.7	931	22	AAU79504	Human protein SEQ
124	308	3.7	1145	22	AGB82169	S. epidermidis ope
125	308	3.7	1312	18	AAU22775	Human RAD50. Homo
126	308	3.7	6281	22	AAU32775	Staphylococcus aur
127	307.5	3.7	795	23	ABB77430	Human tumour marke
128	307	3.6	54	14	AAU41356	MSPIEGF2A EGF2-lik
129	307	3.6	1312	19	AAU71295	Human homologue of
130	306.5	3.6	1211	22	ABB64663	Drosophila melanog
131	306	3.6	1247	21	AAB18215	Plasmodium falcipa
132	304.5	3.6	1182	21	AAB18288	Plasmodium falcipa
133	304	3.6	990	22	AAU78520	Human protein SEQ
134	304	3.6	1398	21	AAB18292	Plasmodium falcipa
135	303.5	3.6	1038	22	AGB67416	Amino acid sequenc
136	303.5	3.6	1975	22	ABB62094	Drosophila melanog
137	303.5	3.6	2434	22	AAU34339	Staphylococcus aur
138	303	3.6	1184	22	AAU33733	Staphylococcus aur
139	303	3.6	1188	22	AAU36532	Staphylococcus aur
140	303	3.6	1192	21	AAB18165	Plasmodium falcipa
141	301	3.6	1046	23	ABB54658	Lactococcus lactis
142	301	3.6	1355	22	ABG00331	Novel human diagno
143	301	3.6	1626	22	ABG20974	Novel human diagno
144	299	3.5	1166	21	AAB18268	Plasmodium falcipa
145	299	3.5	1521	21	AAG39235	Arabidopsis thalia
146	299	3.5	1528	21	AAG39234	Arabidopsis thalia
147	299	3.5	1562	21	AAG39233	Arabidopsis thalia
148	299	3.5	1703	21	AAG36714	Arabidopsis thalia
149	299	3.5	1710	21	AAG36713	Arabidopsis thalia
150	299	3.5	1744	21	AAG36712	Arabidopsis thalia
151	298.5	3.5	2710	17	AAU95016	C. difficile toxin
152	298.5	3.5	2710	19	AAU68387	Clostridium diffic
153	298.5	3.5	3329	21	AAU56029	Human BRCA2 tumour
154	298	3.5	1188	21	AAB18183	Plasmodium falcipa
155	298	3.5	2194	22	AAU40114	Human polypeptide
156	298	3.5	3135	15	AAU57474	P. falciparum tran
157	298	3.5	3135	21	AAB18223	Plasmodium falcipa
158	296.5	3.5	1149	23	ABB77620	AmEPV P4a core pro
159	296	3.5	1663	15	AAU46608	Plasmodium falcipa
160	295	3.5	1588	15	AAU46605	Malarial PfEMP3 ep
161	295	3.5	2637	22	AAU31618	Novel human secret
162	295	3.5	4097	22	ABB57815	Drosophila melanog
163	294.5	3.5	1193	21	AAB18306	Plasmodium falcipa
164	294.5	3.5	1844	21	AAB18250	Plasmodium falcipa
165	294	3.5	1388	23	AAU79590	Human kinesin moto
166	294	3.5	1388	23	AAU79590	Human kinesin supe
167	294	3.5	1786	14	AAU41043	CD4-EBAl75 fuson
168	293.5	3.5	1010	23	ABP40333	Staphylococcus epi
169	293.5	3.5	1084	23	ABB97247	Novel human protei
170	293	3.5	2391	15	AAU56594	Carbamoyl-phosphat
171	293	3.5	3444	22	ABG17495	Novel human diagno
172	292.5	3.5	1123	21	AAU58277	Heliothis armigera
173	292.5	3.5	2025	22	AAU34207	Staphylococcus aur
174	292.5	3.5	2986	22	AAU31620	Novel human secret
175	292	3.5	1341	22	AAU84702	Amino acid sequenc
176	292	3.5	5795	22	AAU37017	Staphylococcus aur
177	291	3.5	1014	23	ABG61923	Prostate cancer-as
178	291	3.5	1327	23	ABG57163	Mouse ischaemic co
179	291	3.5	1960	22	AAU78854	Human protein SEQ
180	291	3.5	2437	22	AAU34338	Staphylococcus aur

ALIGNMENTS

RESULT 1

AAW54145
 ID AAW54145 standard; Protein; 1639 AA.
 XX AC AAW54145;
 XX DT 23-SEP-1998 (first entry)
 XX DE P. falciparum synthetic gp190 protein.
 XX KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 XX OS monoclonal antibody; passive immunisation; parasite.
 XX OS Plasmodium falciparum.
 XX OS Synthetic.
 XX PN WO9814583-A2.
 XX PD 09-APR-1998.
 XX PF 02-OCT-1997; 97WO-EP05441.
 XX PR 02-OCT-1996; 96DE-4040817.
 XX PA (BUJA/) BUJARD H.
 XX FI Bujard H, Pan W, Tolle R;
 XX DR WPI: 1998-240088/21.
 XX DR N-PSDB; RAV21451, RAV35363.
 XX PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
 XX PT protein - useful in anti-malaria vaccines, also stabilising genes by
 XX PS reducing their AT content
 XX PS Example 1; Fig 3c; 48pp; German.

This sequence represents a modified Plasmodium falciparum gp190/MSP-1 (merozoite surface) protein. The gene encoding this protein has been stabilised by reducing the AT content of the nucleotide sequence. Such a protein is useful in vaccines against malaria or for producing monoclonal antibodies (for passive immunisation). The complete gp190 protein can now be produced outside the parasite and has, at least over extended regions, the native pattern of folding. Larger amounts of the protein can be

Db 901 PQDKPVSANDTSHSTNNLSKLFPENILSLCKNKNIIQELIGQSSENTYKILKDS 960

Qy	961	TFYNESGTFNVKSKADINSINLDESKRRKKLEEDINKLAKTKIQLSPDLNKKYKLERLFD	1020
Db	961	TFYNESGTFNVKSKADINSINLDESKRRKKLEEDINKLAKTKIQLSPDLNKKYKLERLFD	1020
Qy	1021	KKETVGCKYKMQIKKLLTKLEQLESKLSLNNPKHHVLQNFSPVFNNKKKAEAEIAETENTLEN	1080
Db	1021	KKETVGCKYKMQIKKLLTKLEQLESKLSLNNPKHHVLQNFSPVFNNKKKAEAEIAETENTLEN	1080
Qy	1081	TKILLKHGKLVKYNGESSPLKTLSEESIQTQEDNYASLENPKVLSKLEGLKKNLNLEK	1140
Db	1081	TKILLKHGKLVKYNGESSPLKTLSEESIQTQEDNYASLENPKVLSKLEGLKKNLNLEK	1140
Qy	1141	KKLSYSSGSLUHLHIAELKEVIKNNKVTGNSPENNTDVNNALLESYKKFLPGDGVATVVS	1200
Db	1141	KKLSYSSGSLUHLHIAELKEVIKNNKVTGNSPENNTDVNNALLESYKKFLPGDGVATVVS	1200
Qy	1201	ESGSDTLTEQSQPKKPASTHVGAESENTITTSQNVDDDEVDDVIIVPITFGSEEDYDDLGQVV	1260
Db	1201	ESGSDTLTEQSQPKKPASTHVGAESENTITTSQNVDDDEVDDVIIVPITFGSEEDYDDLGQVV	1260
Qy	1261	TGEAVTPSVIDNLSIKIENEYEVLYLKPLAGVYRSLKKOLENNVMTFNVNVKDIILNSRFN	1320
Db	1261	TGEAVTPSVIDNLSIKIENEYEVLYLKPLAGVYRSLKKOLENNVMTFNVNVKDIILNSRFN	1320
Qy	1321	KRENFKNVLESDDLIPYKDLTSSNVYVKDYKFLNKEKKRDKFLSSYNYIKDSIDTDINPAN	1380
Db	1321	KRENFKNVLESDDLIPYKDLTSSNVYVKDYKFLNKEKKRDKFLSSYNYIKDSIDTDINPAN	1380
Qy	1381	DVLGYKKILSEKYSKDLSDSIKKYINDKOGENEKYLPLFNNIETLYKTVNDKIDLFVHLE	1440
Db	1381	DVLGYKKILSEKYSKDLSDSIKKYINDKOGENEKYLPLFNNIETLYKTVNDKIDLFVHLE	1440
Qy	1441	AKVLNITYEKSNEYKIKELNYLKTITQDKLADFKKNNNPFVGIADLSTDYNNHNLTKFLS	1500
Db	1441	AKVLNITYEKSNEYKIKELNYLKTITQDKLADFKKNNNPFVGIADLSTDYNNHNLTKFLS	1500
Qy	1501	TGWVFENLAKTVLSNLDGNLQGLMNIISOHQCVKKQCPONSGCFRHLDERECKCLLNYK	1560
Db	1501	TGWVFENLAKTVLSNLDGNLQGLMNIISOHQCVKKQCPONSGCFRHLDERECKCLLNYK	1560
Qy	1561	QEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCBCTKPDSPYPLFDGIFCSSS	1620
Db	1561	QEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCBCTKPDSPYPLFDGIFCSSS	1620
Qy	1621	NFLGISPLLIIMLILYSFI 1639	
Db	1621	NFLGISPLLIIMLILYSFI 1639	
RESULT 2			
AAP50777			
ID	AAP50777 standard; Protein; 1654 AA.		
XX	AAP50777;		
XX	30-SEP-1991 (first entry)		
DE	Sequence of the P195 protein of Plasmodium falciparum.		
XX	Malaria vaccine; epitope; antigen; immunogen.		
KW	Plasmodium falciparum.		
OS	EP154454-A.		
PN	11-SEP-1985.		
XX	21-FEB-1985; 85EP-0301173.		
PR	26-SEP-1984; 84GB-0024340.		
PR	22-FEB-1984; 84GB-0004692.		
PR	21-FEB-1985; 85GB-0004429.		
XX			

Db 61 NFKNVLESLLPYKDLTSSNVVVKPYKFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 120
QY 1384 GYKILSEKYSKDSLSIKKYINDKOGENEKYLPLFLNNIETLYKTVDNKIDLVFIHLEAKV 1443
Db 121 GYKILSEKYSKDSLSIKKYINDKOGENEKYLPLFLNNIETLYKTVDNKIDLVFIHLEAKV 180
QY 1444 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGTADLSTDYNNHNNLLTKFLSTGM 1503
Db 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGTADLSTDYNNHNNLLTKFLSTGM 240
QY 1504 VFNENLAKTVLSNLLDGNLQGMNLISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 1563
Db 241 VFNENLAKTVLSNLLDGNLQGMNLISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 300
QY 1564 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCS 1618
Db 301 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCS 355

RESULT 5

AAV09373

ID AAY09373 standard; Protein; 361 AA.

XX AC AAY09373;

XX DT 31-AUG-1999 (first entry)

XX DE Merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.

XX OS Plasmodium falciparum.

XX PN WO9920774-A2.

XX PD 29-APR-1999.

XX PF 20-OCT-1998; 98WO-US22226.

XX PR 15-MAY-1998; 98US-0085649.

XX PR 20-OCT-1997; 97US-0062592.

XX PA (GENZ) GENZYME TRANSGENICS CORP.

XX PI Chen LH, Meade H;

XX WPI; 1999-288313/24.

XX DR P-PSDB; AAX56009.

XX PT Modified malarial protein for use in anti-malarial vaccines

XX PS Example; Fig 2; 35pp; English.

XX CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing
CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.

XX ' Sequenced 361 AA;

Query Match

Best Local Similarity 22.3%; Score 1881; DB 20; Length 361;

100.0%; Pred. No. 4.5e-75;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1264 AVTPSVIDNLSKINENEYEVLYKPLAGVYRSLLKKOLENNVMTFNVVKDIILNSRFNKE 1323
Db 1 AVTPSVIDNLSKINENEYEVLYKPLAGVYRSLLKKOLENNVMTFNVVKDIILNSRFNKE 60
QY 1324 NFKNVLESLLPYKDLTSSNVVVKPYKFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 1383
Db 61 NFKNVLESLLPYKDLTSSNVVVKPYKFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 120
QY 1384 GYKILSEKYSKDSLSIKKYINDKOGENEKYLPLFLNNIETLYKTVDNKIDLVFIHLEAKV 1443
Db 121 GYKILSEKYSKDSLSIKKYINDKOGENEKYLPLFLNNIETLYKTVDNKIDLVFIHLEAKV 180
QY 1444 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGTADLSTDYNNHNNLLTKFLSTGM 1503
Db 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGTADLSTDYNNHNNLLTKFLSTGM 240
QY 1504 VFNENLAKTVLSNLLDGNLQGMNLISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 1563
Db 241 VFNENLAKTVLSNLLDGNLQGMNLISOHQCVKKQCPQNSGCFRHLDERECKCLLNYKQEG 300
QY 1564 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCS 1618
Db 301 DKCVENPNTCENNGGCCDADAKCTEEDSGSNGKKITCECTKPDSPYPLFDGIFCS 355

RESULT 6

AAV05833

ID AAY05833 standard; Protein; 361 AA.

XX AC AAY05833;

XX DT 02-AUG-1999 (first entry)

XX DE Merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.

XX OS Plasmodium falciparum.

XX PN WO9920766-A2.

XX PD 29-APR-1999.

XX PF 20-OCT-1998; 98WO-US22225.

XX PR 15-MAY-1998; 98US-0085649.

XX PR 20-OCT-1997; 97US-0062592.

XX PA (GENZ) GENZYME TRANSGENICS CORP.

XX PI Chen LH, Meade H;

XX WPI; 1999-302742/25.

XX DR N-PSDB; AAX25587.

XX CC New modified recombinant nucleic acid sequences useful for producing
XX PT malarial DNA vaccine
XX PS Disclosure; Fig 2; 43pp; English.

XX CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
CC AAX25593) encoding MSP-1-42 have been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the

CC 'mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, especially mammalian cell
 CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high AT content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.

XX SQ Sequence 361 AA;

Query Match 22.3%; Score 1881; DB 20; Length 361;
 Best Local Similarity 100.0%; Pred. No. 4.5e-75;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1264 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKR 1323
 Db 1 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKR 60

Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRKDRKFLSSYNYIKDSIDTDINFANDVL 1383
 Db 61 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRKDRKFLSSYNYIKDSIDTDINFANDVL 120

Qy 1384 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 1443
 Db 121 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 180

Qy 1444 LNYTYEKSNEVKIKELNYLKTIOQKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 1503
 Db 181 LNYTYEKSNEVKIKELNYLKTIOQKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 240

Qy 1504 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNYKQEG 1563
 Db 241 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNYKQEG 300

Qy 1564 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 1618
 Db 301 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 355

RESULT 7
 AAY09374
 ID AAY09374 standard; Protein; 376 AA.
 XX AC AAY09374;
 XX DE 31-AUG-1999 (first entry)
 XX DE Modified merozoite surface protein MSP-1-42.
 XX KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.
 XX OS Plasmodium falciparum.
 OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Peptide 1..15
 FT /note= "beta-casein signal peptide"
 FT Peptide 371..376
 FT /note= "6xHis tag"
 FT Misc-difference 197
 FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
 FT Misc-difference 278
 FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
 FT W09920774-A2.
 XX PN 29-APR-1999.
 XX PD
 XX XX

PF 20-OCT-1998; 98WO-US22226.
 XX 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.
 XX (GENZ) GENZYME TRANSGENICS CORP.
 PA Chen LH, Meade H;
 XX WPI; 1999-288313/24.
 DR P-PSDB; AAX56008.
 XX Modified malarial protein for use in anti-malarial vaccines
 PT Example; Fig 11; 35pp; English.
 PS The present sequence represents a 42 kDa C-terminal portion of
 XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The sequence has been modified to include
 CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
 CC tag. In addition, N181Q and N262Q mutations have been introduced to
 CC eliminate 2 N-glycosylation sites. These modifications allow the
 CC MSP-1-42 protein to be expressed in the milk of transgenic mice.

XX SQ Sequence 376 AA;

Query Match 22.2%; Score 1869; DB 20; Length 376;
 Best Local Similarity 99.4%; Pred. No. 1.6e-74;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1264 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKR 1323
 Db 16 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKR 75

Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRKDRKFLSSYNYIKDSIDTDINFANDVL 1383
 Db 76 NFKNVLESDLIPYKDLTSSNVVVDPKYFLNKRKDRKFLSSYNYIKDSIDTDINFANDVL 135

Qy 1384 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 1443
 Db 136 GYYKILSEKYSKDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 195

Qy 1444 LNYTYEKSNEVKIKELNYLKTIOQKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 1503
 Db 196 LNYTYEKSNEVKIKELNYLKTIOQKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 255

Qy 1504 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNYKQEG 1563
 Db 256 VFENLAKTVLSNLDGNLQGLMNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNYKQEG 315

Qy 1564 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 1618
 Db 316 DKCVENPNTCNENGGCDADAKCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 370

RESULT 8
 AAY05834
 ID AAY05834 standard; Protein; 376 AA.
 XX AC AAY05834;
 XX DE 02-AUG-1999 (first entry)
 XX DE Modified merozoite surface protein MSP-1-42.
 XX KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.
 XX OS Plasmodium falciparum.
 OS Synthetic.
 XX XX

Qy 160 NIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELDLVKK 219
|||||
Db 149 NIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELDLVKK 208
|||||
Qy 220 LVFGYRKPLDNIKDNVGMEDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEKKLY 279
|||||
Db 209 LVFGYRKPLDNIKDNVGMEDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEKKLY 268
|||||
Qy 280 QAOYDLSYINKOLEAHNLISVLEKRIIDLTKKNENIKELLOKINEIKNPPPA 331
|||||
Db 269 QAOYDLFIYNKOLEAHNLISVLEKRIIDLTKKNENIKELLOKINEIKNPPPA 320
|||||
RESULT 12
AAP82523
ID AAP82523 standard; protein; 219 AA.
XX
AC AAP82523;
XX
DT 12-DEC-1990 (first entry)
XX
DE 31-1 Repeated Delete.
XX
KW Malaria; stage-specific late schizont-merozoite antigen 31-1;
KW vaccine; protozoan parasite; repitope.
XX
OS Plasmodium falciparum.
XX
PN EP254862-A.
XX
PD 03-FEB-1988.
XX
PF 20-JUN-1987; 87EP-0108867.
XX
PR 26-JUN-1986; 86US-0879076.
XX
PA (BEHW) BEHRINGWERKE AG.
XX
PI Shaw A, Humbert Y;
XX
DR WPI; 1988-030152/05.
DR N-PSDB; AAN82176.
XX
PT Peptide(s) for prodn. of antiprotozoal vaccines - comprising
PT surface antigen sequence with deleted repitope
XX
PS Disclosure; ; p; English.
XX
CC The sequence encodes a polypeptide which contains the unique
CC sequence of the surface antigen of P. falciparum, but lacks the
CC immunodominant repeat sequences which may allow the parasite to
CC evade and decoy the immune system. The DNA for the stage-
CC specific late schizont-merozoite antigen 31-1 in p31-1 (W08503725)
CC was modified to delete the entire sequence coding for the repitope.
CC The expression plasmid, 31-1 Repeated Delete, was introduced into
CC E.coli K12 to produce the modified peptide. The transformant has
CC been deposited in the Deutsche Sammlung von Mikroorganismen as
CC E.coli K12 (9SG936cl) (p31-1 Repeat Delete). The peptide was used
CC in compsns. to treat and prevent protozoan parasitic infections.
XX
SQ Sequence 219 AA;
Query Match 10.8%; Score 910.5; DB 9; Length 219;
Best Local Similarity 76.2%; Pred. No. 8.6e-33;
Matches 182; Conservative 4; Mismatches 8; Indels 45; Gaps 1;
Qy 26 QELVKKLEALEDAVLTYGSLFQEKWVNEGTSFVAVTSTPGSKGVSAGSGGVSAG 85
|||||
Db 6 QELVKKLEALEDAVLTYGSLFQEKWVNEGTSFVAVTSTPGSKGVSAGSGGVSAG 45
|||||
Qy 86 GSVASGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAG 145
|||||

Db 46 -----TNPSDNSSSDSAKSYADLKHVRVONYLTIKELKYP 80
Qy 146 OLFDLTNHLTLCDNHGFYKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPF 205
:|||||
Db 81 ELFDLTNHLTLCDNHGFYKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPF 140
|||||
Qy 206 NLKIRANELDVLKLVFGYRKPLDNIKDNVGMEDYIKKNKKTNIENINELIEESKKTID 264
|||||
Db 141 NLKILRNELDLVKLVFGYRKPLDNIKDNVGMEDYIKKNKKTNIENINELIEESKKTID 199
|||||
RESULT 13
AAP80544
ID AAP80544 standard; protein; 219 AA.
XX
AC AAP80544;
XX
DT 11-NOV-1990 (first entry)
XX
DE Polypeptide p190-1.
XX
KW Polypeptide p190-1; P.falciparum; merozoite; vaccine.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT Region 1..8
FT /label=affinity peptide
FT Region 184..219
XX
XX EP283829-A.
XX
XX 28-SEP-1988.
XX
XX 08-WAR-1988; 88EP-0103564.
XX
XX 19-MAR-1987; 87GB-0006599.
XX
XX (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX
XX Ulrich C, Gentz H, Takacs B;
XX
XX WPI; 1988-272339/39.
XX
XX N-PSDB; AAN81148.
XX
XX New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
XX surface antigen precursor useful in vaccines, and encoding DNA sequences
XX
XX Claim 4; Page 26; 45pp; German.
XX
XX The polypeptide contains at least one epitope of the 190 kD precursor of
XX the major merozoite surface antigen of P.falciparum. The polypeptide is
XX immunogenic and can be used in vaccines to protect against malaria.
XX Presence of the affinity peptide allows the polypeptide to be purified
XX by affinity chromatography resins. Antibodies raised against the
XX polypeptide are useful for diagnosis and for purific. by affinity
XX chromatography.
XX
SQ Sequence 219 AA;
Query Match 10.7%; Score 905; DB 9; Length 219;
Best Local Similarity 98.9%; Pred. No. 1.5e-32;
Matches 174; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 156 TLCDNIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELD 215
|||||
Db 9 TLCDNIHGFKYLIDGYEIEINELLYKLNFFDILLRAKLNDCVANDYCOIPFNKIRANELD 68
|||||
Qy 216 VLKLVFGYRKPLDNIKDNVGMEDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEK 275
|||||
Db 69 VLKLVFGYRKPLDNIKDNVGMEDYIKKNKKTNIENINELIEESKKTIDKNKNATKEEK 128
|||||
Qy 276 KLYQAQYDLSYINKOLEAHNLISVLEKRIIDLTKKNENIKELLOKINEIKNPPPA 331
|||||

Db 129 KKLQAQYDLFIYNKQLEAHNLISVLEKRIIDTLKKNENIKELLDKINEIKNPPPA 184
RESULT 14
AAP50304
ID AAP50304 standard; Protein; 244 AA.
XX AC AAP50304;
XX DT 22-OCT-1991 (first entry)
XX DE Storage-specific, late schizont merozoite malaria antigen.
XX KW Malaria; vaccine.
XX OS Plasmodium spp.
XX PN W08503725-A.
XX PD 29-AUG-1985.
XX PF 20-FEB-1985; 85WO-GB000072.
XX PR 20-FEB-1984; 84GB-0004378.
XX PA (BIOJ) BIOGEN NV.
XX PI Mach B, Perrin L, McGarvey M, Cheung A, Shaw A;
XX DR WPI; 1985-223371/36.
XX DR N-PSDB; AAN50355.
XX PT Prodn. of antigens of Plasmodium species, esp. of falciparum - by
PT recombinant DNA methods giving polypeptide(s) for protecting
PT against malaria or for diagnosis.
XX PS Disclosure; Fig 6; 49pp; English.
XX CC The sequence encodes a Plasmodium falciparum, Plasmodium vivax,
CC Plasmodium malariae and Plasmodium ovale antigen which may be used
CC in the diagnosis of malaria and as a vaccine against malaria.
XX SQ Sequence 244 AA;
Query Match 9.4%; Score 794; DB 6; Length 244;
Best Local Similarity 92.0%; Pred. No. 1.2e-27;
Matches 160; Conservative 3; Mismatches 5; Indels 6; Gaps 2;
QY 1 MKIIFFLCSPFFFIINTQCVTHESYQELVKKLEALEDAVLTYGSLFQEKVMVLNEGTSGT 60
Db 75 MKIIFFLCSPFFFIINTQCVTHESYQELVKKLEALEDAVLTYGSLFQEKVMVLNEGTSGT 134
QY 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGSGNSRRTNPSDSSDS 120
Db 135 AVTTSTPGSGSVTSGGSGGVA---SVASGG---SGGVSAGSGGNSRRTNPSDSSDS 188
QY 121 DAKSYADLKHRVNYLTIKELKYPOLFDLTNNHMLTCLDNIHGFYLDGYEEI 174
Db 189 DAKSYADLKHRVNYLTIKELKYPOLFDLTNNHMLTCLDNIHGFYLDGYEEI 242
RESULT 15
AAP80546
ID AAP80546 standard; protein; 151 AA.
XX AC AAP80546;
XX DT 11-NOV-1990 (first entry)
XX DE Polypeptide p190-2b.
XX KW Polypeptide p190-2b; P.falciparum; merozoite; vaccine.

XX Plasmodium falciparum.
XX FH Key Location/Qualifiers
FT Region 1..5
FT FT /label-affinity peptide
FT FT 142..142
XX PN EP283829-A.
XX PD 28-SEP-1988.
XX PF 08-MAR-1988; 88EP-0103564.
XX PR 19-MAR-1987; 87GB-0006599.
XX PA (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX PI Ulrich C, Gentz H, Takacs B;
XX DR WPI; 1988-272339/39.
XX DR N-PSDB; AAN81150.
XX PT New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
PT surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX PS Claim 6; Page 27; 45pp; German.
XX CC The polypeptide contains at least one epitope of the 190 kD precursor of
CC the major merozoite surface antigen of P.falciparum. The polypeptide is
CC immunogenic and can be used in vaccines to protect against malaria.
CC Presence of the affinity peptide allows the polypeptide to be purified
CC by affinity chromatography resins. Antibodies raised against the
CC polypeptide are useful for diagnosis and for purificn. by affinity
CC chromatography.
XX SQ Sequence 151 AA;
Query Match 8.0%; Score 677; DB 9; Length 151;
Best Local Similarity 95.8%; Pred. No. 9.5e-23;
Matches 136; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1069 AEIAETENTLENTKILLKHVKYNGESSPLKTLSEESTQTDNYASLENFKVLSKL 1128
Db 6 AEIAETENTLENTKILLKHVKYNGESSPLKTLSEESTQTDNYASLENFKVLSKL 65
QY 1129 EGKLDNLDLLEKKLSYLSGLHLLHLLAEIKVKNKNTGNSPSNNTDVNNALSYKKF 1188
Db 66 EGKLDNLDLLEKKLSYLSGLHLLHLLAEIKVKNKNTGNSPSNNTDVNNALSYKKF 125
QY 1189 LPEGTDVATVVSSESGSDTLEQS 1210
Db 126 LPEGTDVATVVSSESGSDVLDQPS 147
RESULT 16
AAW36103
ID AAW36103 standard; Protein; 116 AA.
XX AC AAW36103;
XX DT 25-MAR-1998 (first entry)
XX DE PFMSP1(p19)A protein sequence.
XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Region 1..95

FT /note= "amino acids derived from P. falciparum MSP1 p19
FT 96.116
FT /note= "glycosylphosphatidylinositol anchoring sequence"
XX
XX
PN WO9730158-A2.
XX
XX PD 21-AUG-1997.
XX
XX PF 14-FEB-1997; 97WO-FR00291.
XX
XX PR 14-FEB-1996; 96FR-0001822.
XX
XX PA (INSP) INST PASTEUR.
XX PA (UYNV) UNIV NEW YORK STATE.
XX
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX
XX DR WPI; 1997-425033/39.
XX DR N-PSDB; AAT94550.
XX
XX PT Recombinant protein containing the merozoite surface protein-1 p19
PT fragment - useful in anti-malarial vaccines, diagnosis and protein
PT purification
XX
XX PS Disclosure; Fig 1B; 85pp; French.
XX
XX This is the amino acid sequence of a recombinant protein comprising
CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX SQ Sequence 116 AA;
Query Match 7.6%; Score 637; DB 18; Length 116;
Best Local Similarity 99.1%; Pred. No. 3.9e-21;
Matches 113; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1526 NISQHCYKVKQCPQNSGCFRHLDERECKCLLNLYKQEGDKCVENPNTCNENNGGDADA 1585
Db 3 NISQHCYKVKQCPQNSGCFRHLDERECKCLLNLYKQEGDKCVENPNTCNENNGGDADA 62
QY 1586 KCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFGISFLLILMLILYSFI 1639
Db 63 KCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFGISFLLILMLILYSFI 116
RESULT 17
AAW22592
ID AAW22592 standard; Protein; 116 AA.
XX
XX AC AAW22592;
XX
XX DT 25-MAR-1998 (first entry)
XX
XX DE PfMSP1(p19)A protein sequence.
XX
XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX Region 1..95
FT /note= "amino acids derived from P. falciparum MSP1 p19
FT fragment"
FT 96.116
FT Region

FT /note= "glycosylphosphatidylinositol anchoring sequence"
XX
XX PN WO9730159-A2.
XX
XX PD 21-AUG-1997.
XX
XX PF 14-FEB-1997; 97WO-FR00291.
XX
XX PR 14-FEB-1996; 96FR-0001821.
XX
XX PA (INSP) INST PASTEUR.
XX PA (UYNV) UNIV NEW YORK STATE.
XX
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX
XX DR WPI; 1997-425034/39.
XX DR P-PSDB; AAW22592.
XX
XX PT Recombinant protein containing Plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX
XX PS Disclosure; Fig 1B; 85pp; French.
XX
XX This is the amino acid sequence of a recombinant protein comprising
CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX SQ Sequence 116 AA;
Query Match 7.6%; Score 637; DB 18; Length 116;
Best Local Similarity 99.1%; Pred. No. 3.9e-21;
Matches 113; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1526 NISQHCYKVKQCPQNSGCFRHLDERECKCLLNLYKQEGDKCVENPNTCNENNGGDADA 1585
Db 3 NISQHCYKVKQCPQNSGCFRHLDERECKCLLNLYKQEGDKCVENPNTCNENNGGDADA 62
QY 1586 KCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFGISFLLILMLILYSFI 1639
Db 63 KCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFGISFLLILMLILYSFI 116
RESULT 18
AAW50527
ID AAW50527 standard; Protein; 115 AA.
XX
XX AC AAW50527;
XX
XX DT 12-MAR-2002 (first entry)
XX
XX DE Plasmodium falciparum MSP-1 peptide p115MSP-1.
XX
XX KW Merozoite surface antigen 1; MSP-1; p115MSP-1; antigen; immunogen;
KW malaria; vaccine; Venezuela equine encephalitis virus; DVEE;
XX vector; immunisation.
XX
XX OS Plasmodium falciparum.
XX
XX PN WO200185927-A1.
XX
XX PD 15-NOV-2001.
XX
XX PF 08-MAY-2001; 2001WO-US14716.
XX
XX PR 08-MAY-2000; 2000US-202430P.
XX

PA (GEOU) UNIV GEORGETOWN.

XX
PI Davidson E, Nikodem D;
XX
XX WPI; 2002-049444/06.
XX
XX Novel vaccine for immunizing mammals against Plasmodium falciparum
PT infection, comprises a viral vector system expressing protein
PT corresponding to specific domain of major merozoite surface protein 1
PT of Plasmodium falciparum
XX
XX Disclosure; Page 4; 60pp; English.

XX The present sequence is that of immunogenic protein p115MSP-1, a
CC specific domain (amino acids 1002-1116) of the major merozoite
CC surface protein 1 (MSP-1) of the erythrocyte stage of the
CC West African Wellcome strain of Plasmodium falciparum. The
CC invention relates to a malaria vaccine comprising an expression
CC vector, preferably a defective Venezuelan equine encephalitis (DVEE)
CC viral vector system, which expresses p115MSP-1 or its immunogenic
CC fragment after administration to a patient. The p115MSP-1 protein
CC or fragment raises a humoral and/or cell-mediated response to the
CC erythrocytic merozoite malaria antigen, protecting the patient from
CC a subsequent malaria infection. The DVEE viral vector system
CC continues to express antigen in the patient for a period of days,
CC months or even years. The p115MSP-1 antigen may also include
CC a signal peptide and/or an anchor peptide sequence, which
CC enhances the immunogenic response of the patient.

XX Sequence 115 AA;

Query Match 6.9%; Score 584; DB 23; Length 115;
Best Local Similarity 100.0%; Pred. No. 8.1e-19;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1002 QLSFDLYNKYKLERLFDKKTGKYKMQIKKTLLEQLESKLSLNPKHVLQNFVSV 1061
Db 1 QLSFDLYNKYKLERLFDKKTGKYKMQIKKTLLEQLESKLSLNPKHVLQNFVSV 60
Qy 1062 FFKKKKEAIEAETENTLENTKILLKHYKGLVKYNGESSPLKTLSEESIQTEDNY 1116
Db 61 FFKKKKEAIEAETENTLENTKILLKHYKGLVKYNGESSPLKTLSEESIQTEDNY 115

RESULT 19
AAB37609
ID AAB37609 standard; Protein; 108 AA.
XX
AC AAB37609;
XX

27-FEB-2001 (first entry)
XX
XX Merozoite surface protein-119.
XX
XX Merozoite surface protein; protazoacide; vaccine; malaria.

Plasmodium falciparum.

WO2000063245-A2.

26-OCT-2000.

20-APR-2000; 2000WO-GB01558.

20-APR-1999; 99GB-0009072.

13-MAY-1999; 99US-0311817.

25-MAY-1999; 99CA-2271451.

(MEDI-) MEDICAL RES COUNCIL.

Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiipbull C;

WPI; 2001-015762/02.

DR N-PSDB; AAC68977.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria
XX

XX Example 5; Fig 15; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is MSP-119 protein.

XX Sequence 108 AA;

Query Match 6.6%; Score 560; DB 22; Length 108;
Best Local Similarity 94.1%; Pred. No. 8.5e-18;
Matches 96; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1520 NLQGLMNTSQHCVKVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNG 1579
Db 7 HIEGRHNIAHQCVKVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNG 66

Qy 1580 GCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSN 1621

Db 67 GCDADAKCTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSN 108

RESULT 20

AAB18171
ID AAB18171 standard; Protein; 1979 AA.

XX AC AAB18171;

07-NOV-2000 (first entry)

Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.

Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
antimalarial; malaria; protozoacide; infection; insecticide.

Plasmodium falciparum.

WO2000025728-A2.

11-MAY-2000.

05-NOV-1999; 99WO-US26796.

05-NOV-1998; 98US-0107131.

(HOFF/) HOFFMAN S.

(CARU/) CARUCCI D.

(GARD/) GARDNER M.

(VENT/) VENTER J C.

Hoffman S, Carucci D, Gardner M, Venter JC;

WPI; 2000-365347/31.

Proteins encoded by chromosome 2 of the human malarial parasite,
Plasmodium falciparum, useful as antimalarial vaccines and in the
diagnosis of P.falciparum infection -

Disclosure; Page 70-75; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against *P. falciparum* infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with *P. falciparum*. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent *P. falciparum*
 CC infection, or they can be used to identify drug resistance in
 CC *P. falciparum*. Sequencing of the plasmid chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasite lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AA70078 to AA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.

XX SQ Sequence 1979 AA;

Query Match 6.6%; Score 559.5; DB:21; Length 1979;
 Best Local Similarity 22.2%; Pred. No. 2.9e-16;
 Matches 370; Conservative 314; Mismatches 609; Indels 373; Gaps 82;

QY 110 RTNPSDNSSDAKSYADLKHVRNYLTITIKELYPQLFDLTFNMLTLCDDNTHGPKYLLD 169
 DB 126 KNELENQLKDLTKSSLSNKNVYNVESKIEBE-KELKEVKDKNI---DN-----N 173
 QY 170 GYEENELLYKLNFY---FDLLRAKLVNDVCANDYQOIFPNLAKIRANELDVLKLVGGRK 226
 DB 174 DYE--NKLKEDEKDFVKQIDMLNKEN-----LQEKELDKINKR-----EK 212
 QY 227 PLDNTKDNVKGME-----DYIKNNKKTININELTBESKTTIDKNNKNAKKEBKKL 278
 DB 213 KINEKEKNIKEEFPNHEKYELENKER-ETISIEIIDIKHLEKLEKIEKEKE--- 268
 QY 279 YQAOYDLSYNNKOLBEAHLISVLKRIKIDTLKKNENIKELDKINEIKN-----PPANS 333
 DB 269 -----DLENLKKLSKENVLKEKGVK--EKNETINSLNDNIETEKEKKYKLLVELEE 321
 QY 334 GNPWTLLDKNKKIEBEKE---IKELAKTIFNDSLFTDPLELEYLYREKNNKIDIS 389
 DB 322 KKKQIDLLNKKQBEKEKEKEKEKEKEKEKEKEKEYDTLIKELKDEKISLEKVVHSIKVRE 381
 QY 390 AKVETKESTEPENYNGVTPYPLSYNDINNALNEL-NSFGDLINPFDTYKEPSKNITWTNE 448
 DB 382 MDIEKREH-----NFLHMDQLKDLNSFYKNNQLKVKYCEIKNLKTELE 427
 QY 449 RK-----KFINIEKEKIKIEKKIESDKSYEDRSKSLNDITKEYEKLL 492
 DB 428 KKEKELKDIEVSKESKEINKLINQNEK-----EKQILAFNKNHKEETHGLKEELKESVK-I 482
 QY 493 NEIYDSKFNNDLNFPE-KMMGKRYSYKVEKLTHINTFASYENSKHNLEKLTALKYME 551
 DB 483 TKIETQELQEMVDIKOKELDQLOERYNAQIESI-----SIELSKKEYNQYKNYITE 535
 QY 552 DYSLRNIVVEKELKYKNL-----ISKIENEIEFTLVENIK-----KDEOLPEK 595
 DB 536 EINNLEKELEETNKETYNLQNNYTNELNNDIHLNGLNITMTQISTLANDVHLLNE 595
 QY 596 KITKDENKP---DEKILEVS-DIVKVOQVLLMKRIDEL-KKTQILKKNVELKHN--TH 648
 DB 596 QIDKLNNEKGTLSKISLNVQIMDLKKEKDFLNQIVDLSNQIDLLTRKMEKEKNKMLE 655
 QY 649 VPNSYKQE-----NKQEPYIYLVKKEIDKLVKMPKVESLINEKKNIKTEGQSDNSE 702
 DB 656 QENKYKQEMELLRGNIKSSSENILNNDDEYCDLK-----RKLSLSESEMKMKKEHDKKLA 711
 QY 703 PSTEGEITGQATTKPQQAQSALE---GDSVQAQAEQKQAPFPVPVPEAKAQPVTTP 759
 DB 712 LKDDCQVRIREMNEKNEDKINMLKEEYEDKINTLKEQNEED----- 751

QY 760 APVNN-KTENYSKLDYLEKLYEF-LNT---SYICHKYILVSH-----STMNEKILKQYKI 809
 DB 752 -KINTLKEQNEDKINTLKEEYEHKINTMKEEYEHKINTLNEQNEHKINTLNEQNEHKINT 810
 QY 810 TKEE-ESKLSGCDPLDLFNQNTNIPVYMPDSLNLSQLFM--EYEREMVCLNLYKL 866
 DB 811 MKEEYEDKMTLN-----EQNEDKMNSLKEEYENKQINSNNEIKIDVV-NEY-I 860
 QY 867 KDNKIKNLLKEAK---VSTSVKTLSSSMOPLSLTPQDKPEVSANDDTSTHNLNSL 923
 DB 861 EVDKLVTLDEKKKOFKOFKINAHKAHEKEQIILTEME--ELACQDNKYSDILYKYI 918
 QY 924 KLFENILS-----LGKNNKIYOELIGQKSSNFYKILKDSOT 961
 DB 919 KLIKSIICMIINIECCDDIENEDIRRIEYINNKGAKKEVEKEKHKHSSNILKSEK 978
 QY 962 FYNESFTNFVSKADDINSLNDE---SKRKLLEEDINKLKTQLSLDFLYNLYKLKLEL 1018
 DB 979 F-----FKNSIEDKSHELKKHEKDLKSKDEIEE---NNKKIKELNNDI---KKLQDEIL 1028
 QY 1019 FDKKKTGVKYMQIKKLTLL-----KEQLESKLNSLNNPKHVLFQNFVFFNKKKEAEAE 1073
 DB 1029 VYKQSNQAQVDHKKKSWILLKDKSKKIKDKENQINVEKNEEDL-----KKKDEI-- 1081
 QY 1074 TENTLENTKILLKHVGLVYK---YNGESSPLKTLSESIQOTEDNYASLENFKVLKLE 1129
 DB 1082 -----RILNEE---LVKYKTYILNKLKDPQLQNDLLSKIDINSLTINBGMCDKIE 1130
 QY 1130 GKLD---NLNLEKKLSYLSGLHLHJAELEKVIKKNYTGNSPSEN-----TD 1177
 DB 1131 EHILDYDEINKSRNLFQKNEICSLTTEYVWELANKN---ELIEENKMLVDQGGKK 1187
 QY 1178 VNNALESYKFLPEGTVDVATVYVSESGDTLEQSPKPKPASTHVGAESNTITTSQNVDDV 1237
 DB 1188 LKKDVEKQKKEIKELNKQLTCKNQIDELNEEVEKLNENIELITYSNDL---NNKFDMK 1244
 QY 1238 DVIIVPIFGESEEDYDLGQVVTGEAVTPSIDNLSKIEN-EYEVLY---LAPLAGV 1292
 DB 1245 ENLMMKL---DENEDN-----IKMKSKIDDMEKEIKYREDEKKNLNE 1286
 QY 1293 YRSKAKQLENVMTFN-VNVK---DILNSRFNKNRFKNV-LESDLTPYK-DLTSSNVYVK 1347
 DB 1287 INNLAKKNEKDMCIKYNENNIKYGDCVKEEMLTYKETSLEYEQIKVYDEKCSQY--- 1343
 QY 1348 DPKYFLNKEKDKF-----LSSYNYIKDSIDTFINFAVDVILGYKILSEKYSKSLD 1398
 DB 1344 DEIRFOYDEKCFQYDEINKKYGALLNITNKNWDSKVDRNNN-----EIIISVDNK-- 1396
 QY 1399 SIKKINDKQGENEKYLPF---LNNIETLYKT-VNDK---IDLFIHLEAKVLNYTYEKN 1452
 DB 1397 GIANYLKQIFELNEEIRLKGINKISLISNELNEKNSYDINMKHIOPLL--FLEKTN 1454
 QY 1453 VEVKIKELNYLTQIDKLADFKNNN-----FVGIADLSTDYNNHNNLLTKFLSTGMV 1504
 DB 1455 KENEKIINTSQYSDA---YKKSDSKLCAQFVDDVNIYGNISNNNIRNEYKYEEM 1511
 QY 1505 FENLAKTVLSNLDGLOGMLNISQCVKQCPQNSGCFRHLDERECKC-----LLN 1558
 DB 1512 FD-----TNIEKN--GM-HLSKVIHLEB--NKFRCKMKIYENENIKSSNKIIGLYN 1559
 QY 1559 YKQ---EGDKC-----VENPNTCTENNGGCCDA-DAKCT 1588
 DB 1560 YSRYTGLRDLCKEEIVPSKIGNISKNKNNNKNNKNTCDGYDEKVT 1605

RESULT 21

AAB37608

ID AAB37608 standard; protein; 96 AA.

XX AAB37608;

XX 27-FEB-2001 (first entry)

XX Merozoite surface protein-1.
DE Merozoite surface protein; protozoacide; vaccine; malaria.
KW Merozoite surface protein; protozoacide; vaccine; malaria.
XX Plasmodium falciparum.
OS WO200063245-A2.
XX 26-OCT-2000.
XX 20-APR-2000; 2000WO-GB01558.
XX 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX (MEDI-) MEDICAL RES COUNCIL.
XX Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaiipibull C;
XX WPI; 2001-015762/02.
XX Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria.
XX
XX Example 2; Page 48; 126pp; English.
XX The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The present sequence is the
CC wild-type MSP-1 protein. This sequence was used to generate the variants
CC of the present invention. The non-natural variants of the present
CC invention are useful for immunising a mammal against malaria, and can be
CC used to treat malaria.
XX
XX Sequence 96 AA;
Query Match 6.6%; Score 556; DB 22; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.1e-17;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1526 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 1585
Db 1 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 60
QY 1586 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSN 1621
Db 61 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCSSN 96
RESULT 22
AAW22593
ID AAW22593 standard; Protein; 127 AA.
XX AAW22593;
XX 25-MAR-1998 (first entry)
XX PFMSPl(p19)S protein sequence.
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX Location/Qualifiers
FH

Peptide 1..19
FT /note= "signal peptide"
FT Protein 20..127
FT /note= "mature protein"
FT Region 1..32
FT /note= "derived from P. vivax MSP1"
FT Region 33..34
FT /note= "encoded by restriction enzyme sequence used to
FT create the chimeric sequence"
FT Region 35..127
FT /note= "derived from P. falciparum C-terminal p19
FT fragment of MSP1"
XX WO9730159-A2.
XX 21-AUG-1997.
XX 14-FEB-1997; 97WO-FR00291.
XX 14-FEB-1996; 96FR-0001821.
XX (INSP) INST PASTEUR.
XX (UJNY) UNIV NEW YORK STATE.
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX WPI; 1997-425034/39.
XX P-PSDB; AAW22592.
XX Recombinant protein containing plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX This is the amino acid sequence of a chimeric protein comprising amino
CC acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
CC linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
CC C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 127 AA;
Query Match 6.4%; Score 539; DB 18; Length 127;
Best Local Similarity 98.9%; Pred. No. 8.6e-17;
Matches 92; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1526 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 1585
Db 35 NISQHCVKKQCPQNSGCFRHLDERECKLLNKKQSGDKCVENPNTCNENGGCDADA 94
QY 1586 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 1618
Db 95 KCTEEDSGSGNGKKITCECTKPDSPYPLFDGIFCS 127
RESULT 23
AAW36102
ID AAW36102 standard; Protein; 127 AA.
XX AAW36102;
XX 25-MAR-1998 (first entry)
XX PFMSPl(p19)S protein sequence.
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX


```
QY 536 -----SKHN-----LEKL-----TKALK----- 548
DB 495 TEKNRLNGIDSTITNIEGALKRESKNGYEIGFLEKLEEGKRNKLVDTIKKSINSITVGNF 554
QY 549 --YMEDYSLRNIIVPEKELKYKNLSKTEJIE-----TLVENIKDEBOLFEKTKTKDEN 602
DB 555 SSLENNFOLNOYDFENKNINDYENKNGEYINFEFGLNRISENLRNASEN-----TSDYN 608
QY 603 KPDEKILEVSDIVKQVQKVLMMKNKIDELKKTQLTLKLVKELKHNHVPNSYKQENKQEPY 662
DB 609 SAKTURL-----AQEKVNLNKEEANKYLRDVKVE-----SPR----- 645
QY 663 YLIVLKEIDKLKVPMPKVESLINEEKNKIK-----TEQSDNS 701
DB 646 FIFNMKESLDKINEMIKKEQLTVNGBGVNQLVENIKELVDENNLSDILKAQATGKNEE 705
QY 702 EPSTEGETGQATKFGQOAGSA-----LEGDSVOAQAOBOKQAOQPPVPV 746
DB 706 QKITHSTLKNKAKTILGHVDTSAKYVGKITPELALTELLGDAKLTAQELKF----- 758
QY 747 PVPEAKAQVPTPPAPVNN---KTEVSKLDVLEKLYEFLNYSYICHKYLIVSHSTMNEKI 803
DB 759 ---ESK-----NNVLETENSK-----TNEL-----DVHKN 784
QY 804 LKQYKITEESKSSCDPLDLLEFNIONNIPVMYSMFDSLNNLSOLFMEIYKEMVCNL 863
DB 785 QDAYKVALE---ILAHSEIDT-----KOKDS---SKLIEMGNQIYLVKVLINQ 827
QY 864 YKLKNDKIKNLEBAKVSYSVTKTSSSSMQPLSLTPQDKPEVSANDDTSHSTNLSL 923
DB 828 YK---NKISSIKKEEAVSVKIGNVSKKHSLSKITSKDS----- 865
QY 924 KLFENILSLGN---KNYIQELIGOKSSENFYEKTLKDSDFYNESFTNFVSKRADDINS 980
DB 866 --YDNIILEKQTELQNLNRSFTQKTNDSKLEKIKTFE---ESLKNALKTLEGEVNA 921
QY 981 L-----NDESKKKLEEDINKKTLQLSFDLYNKKYKILKLERLDFKKT---VGRYKMOIK 1033
DB 922 LKASSDNHEHVQSK-SEVPNPALSEIEKEETDIDSLNTALDELLKGRTECVSRVK----- 976
QY 1034 KLTLLKEQLESKLSLNNPKHVLFQNFVFFNKKKEAEAEFTENTLE-NTKLILKHYKGLV 1092
DB 977 ---LIKOTV-----TKEISDDTELINTIEKNVKAYLAIK--- 1008
QY 1093 KYNGESSPLKLTSESIQTEDNVASLENFKVLSKLECKLDNLEKKLSYLSGLLH 1152
DB 1009 KNYEDTVQDVLTLNEHF-----NTQVSNHEPTNFDKSNKSEELTKAVTDSKT 1057
QY 1153 LIAELKEVKNKNYTGNSPSENNTDVNNALSEYKFKFLPEGTDAVTVVSESGSDTLEQSQP 1212
DB 1058 IISKLKGVIEVN-----ENTEM-NTIESSAK-----EIEALYNE-----LKN 1094
QY 1213 KKPASTHVGAESNITTTSQ---NVDEVDVDDVLIPIFGESEEDYDDLQGVVTVGEAVTPSV 1269
DB 1095 KKTSLNEIYQTSNEVKYLOEMKSNADKYID----- 1123
QY 1270 IDNLSKITEYEVLYLKLPLAGVYRSKLENNVMTFNVVKNKDLNSRENKFNKVL 1329
DB 1124 ---VSKIFNT-----VLDTQKSNIVTNOHSIN-NVKDKLKGK-----LQELI 1161
QY 1330 ESDLIPIKDLTSSVNVKDPYKFLNKKRDRKFLSSYNYIKDSIDTIDINFANDVLGYKIL 1389
DB 1162 DAD-----SSFTL-----ESIKKEFNEYSHIKTNI-----GELEQL 1192
QY 1390 SEKYSKLDLSTKKYNDKQGENEYKLPPLNNIETLYKTV-NDKIDLFIHLEAKVLN 1445
DB 1193 QOTNKSCHDNVAKH-----KEXIVHLINRVESLKGDKVKNHDDQYMKKLNASILN 1242

RESULT 25
AAW24575
ID AAW24575 standard; Protein; 1254 AA.
```

```
XX
AC AAW24575;
DT 10-NOV-1997 (first entry)
DE Merozoite apical-end protein clone 5.3.
KW Merozoite apical-end protein; MAEP; Plasmodium vivax; antigen; malaria;
KW human; erythrocyte; antimalarial vaccine; antibody; blood stage parasite;
KW Duffy blood group antigen; red blood cell; therapy.
OS Plasmodium vivax.
PN US5646247-A.
PD 08-JUL-1997.
XX 05-APR-1989; 89US-0334041.
XX 04-OCT-1991; 91US-0792865.
XX 05-APR-1989; 89US-0334041.
PR 06-APR-1989; 89US-0334270.
PR 03-APR-1990; 90WO-US01849.
PR 02-NOV-1990; 90US-0608639.
XX (UJNY ) UNIV NEW YORK STATE.
PA Barnwell JW, Galinski MR;
PI WPI; 1997-362995/33.
DR N-PSDB; AAT80072.
XX Plasmodium merozoite apical end protein - useful as antigen for
PT production of anti-malarial vaccines
PS Claim 1; Column 29-38; 68pp; English.
XX
CC AAW24575 and AAW24576 represent the merozoite apical end proteins (MAEP)
CC isolated from two different Plasmodium vivax strains. These proteins are
CC the antigens of the invention, and immunoreact with antibodies against a
CC native MAEP sequence. P. vivax is one of the four malarial species that
CC infects humans, and is difficult to target for a vaccine, as it cannot
CC be cultured in vitro. The preinvasion orientation of malarial merozoites
CC indicates that the apical end plays an important role in the invasion
CC process. The MAEP protein binds to the surface of susceptible
CC erythrocytes from P. vivax susceptible humans and primates, and also
CC binds to rabbit erythrocytes. The antigen can be used for the production
CC of antimalarial vaccines. The antigens are involved in the invasion
CC process, and are immunochemically reactive with antibodies raised
CC against malaria (particularly P. vivax) blood stage parasites. Synthetic
CC proteins, polypeptides, peptide fragments and analogues of these
CC antigens can be used similarly. As the antigens specifically bind to a
CC Duffy blood group antigen (the antigen present on the surface of
CC susceptible mammalian red blood cells), and are necessary in the process
CC of invasion of red blood cells by merozoites, they can be used to inhibit
CC the invasion of red blood cells by a malarial organism. The antigens can
CC also be used in a method for inhibiting invasion of susceptible
CC mammalian blood cells by malarial merozoites, and in a method for
CC inhibiting the propagation of a malarial organism in susceptible red
CC blood cells.
SQ Sequence 1254 AA;
Query Match 5.4%; Score 455.5; DB 18; Length 1254;
Best Local Similarity 20.5%; Pred. No. 6.1e-12;
Matches 319; Conservative 221; Mismatches 469; Indels 548; Gaps 71;
QY 115 DNSSDSDAKSVADLKHVRVNYLLTIKEL-----KYPQLFDLTNHLMT 156
DB 8 DTSEDEKKKSIEKAYEKMGN---TLKELEKMDDEKNEKEVEEAEIQYKRIF--IDHVDN 62
QY 157 LC-DNIHGFYKYLIDGYE----EINELLYKLNFYFDLLRAKLNDCANDYCOIPFNLKIRA 211
```


Db 63 LMNDEVEKSIWMEKIELYKKEIDEIKOKTNEY-----KQGD-T-SNFY----- 104
QY 212 NELDVLKLVGYRKLPLDNKONVGKMDYI-----KKNKXTIENINELIBESKTTDK 265
Db 105 -----YTQYNSATOSKAKIEQFINIATTKGTSDTSODINEL-ESIEEYHK 151
QY 266 KKNATKE-----EKKKLYQAOYDSIYNKQ-----LEAHNLISVLEKRIDPLKKNENIK 316
Db 152 NLQLVKQESNMEEMRKQILSKMDLLILNNSETIAKEISNTQNALGFRENAKTKLNKTD 211
QY 317 ELLDKINEKKNPPANSNPWTILDKN-----KKIEHEKEIKEIAKTIKFNIDSL--F 369
Db 212 ELLORAAAMIEBAKAKNNIDIALEDAQIDTEVSIEQINRMKDKDEIKSYLSEIKY 271
QY 370 TDPLEYVLRKKNIDISAKVETKESTEPNEYNGVTPYPLSYNDINNALNELSFGDL 429
Db 272 KDKCTEISNRGKD-----KIEPLEKFPNEENS-----NKNV--INEINE--NI 315
QY 430 INPFOYTRK-----PSKNIYTDNE-----RKFFIN--EIREK 459
Db 316 RNSEQYLDIEDAEKQASTKVELFHKHETTISNIFKESIELGVETKSKQKINKAEDIMKE 375
QY 460 IKIEKKTESDKSKYEDSKSLNDITKEYEKLNLIELYDSKPN-----NI- 504
Db 376 IERHNSIEQTQVKGQENLKNLE--PHNYDNAEDELNDKSTNAKVLIETNLESYKHNLS 434
QY 505 DLTNPEKMMGKRYSYKVE-----KLTHNT-----PASYEN----- 535
Db 435 EITNIKQGEKIYSKADIMQIKATSENTAEKTLKVKDDQSNVYVNLQITTERNLIV 494
QY 536 -----SKHN-----LEKL-----TKALK----- 548
Db 495 TEKNRLNGIDSTITNEGALKESKNGYEIGFLEKLEEGCKNKKLVKITKKSINSTVGNF 554
QY 549 ---YMEDYSRLNIVWEKELYKNLISKTEJETE-----TLVENIKKDEQLPEKKTITKDBN 602
Db 555 SSLFNFDLQYDFNKNIDYENKGEIYNEFEGLSKISENLNASEN-----TSYN 608
QY 603 KPDEKILEVSDIVKQVQKVLMLNKIDELKTKTOLILKNVLEKHNIHVPSYKQENKQEPY 662
Db 609 SAKTLRLE-----AQEKVNLNKEEAEANKYLRDVKVE-----SFR----- 645
QY 663 YLIVLKKEIDKLVKMPKVESLINEKKNIK-----TEGOSONS 701
Db 646 FIFNMKESLDKINEMIKKQLFVNEGHVGNVKNOLVENIKELVDENNLSDLTKQATGKNBEI 705
QY 702 EPSTEGETGOATTRPGQOAGSA-----LEGDSVQAOAQEQKQAPPPVPV 746
Db 706 QKITHSTLKNKAKTILGHVDTSKAYGVKITPELALTELLGDAKLKTAQELKF----- 758
QY 747 PVPKAAQVPTPPAPVNN---KTENVSKLDYLEKLYEFLNTSYICHKYLVSHTMNEKI 803
Db 759 ---ESK-----NNVLETENNSK-----NTNEL-----DVHKNI 784
QY 804 LKQYKITKEESKLSGCDPLDLFIQNNIPWYVSNFOSLNSLSQLFMEIYEKEWCVNL 863
Db 785 QDAYKVALE---ILAHSDSIDT-----KOKDS--SKLIEMGNOIYLVKVVLIQ 827
QY 864 YKLKNDKIKNLLEAKKYSTSVKTLSSSSMQPLSLTQDKPKEVSAANDTSHSTLNLSL 923
Db 828 YK-----NKISSKSEAEVSVKIGNVSKHSELSKITCDKS----- 865
QY 924 KLFNILSLGKN---KNIQOELIGOKSSENFYKILKDSDTFYNESFTNVFKSKADDINS 980
Db 866 ---YDNIIALEKOTELONLNSRSTQKNTWNSDKLEKIKTDF--ESLKNALKTLEGEVNA 921
QY 981 L-----NDESKRKKLEEDINKLTKTLQISFDLYNNKYKLEKLFDPKKT--VGKYMOKIT 1033
Db 922 LKASSDNHERVQSK--SEPVNPALSETEKEETDIDSLNTALDELLKKGRTCEVSRVK---- 976
QY 1034 KLTLKLEQLESKLSNLNPKPHVLQNFVFPFNKKKAEIETENTLE--WTKILLKHVKGLV 1092
Db 977 ----LIKDTV-----TKISDSDTELTINTIEKNVKAIVAYIK---- 1008

RESULT 26

AAB18172
ID AAB18172 standard; Protein; 2485 AA.
XX
AC AAB18172;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:29.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite.
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 75-82; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are refined or secreted or membrane proteins)

Staphylococcus epidermidis SRI strain; infection; diagnosis;
vaccination; endocarditis.

Staphylococcus epidermidis.

WO200134809-A2.

17-MAY-2001.

09-NOV-2000; 2000WO-US30782.

09-NOV-1999; 99US-0164258.

(GLAX) GLAXO GROUP LTD.

Kimmerly WJ;

WPI; 2001-316495/33.

N-PSDB; AAH53785.

Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
useful for vaccinating against infections, e.g. endocarditis -

Claim 18; Page 779-781; 2188pp; English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the S. epidermidis polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to AAH55090 represent specifically claimed S. epidermidis genomic DNA polynucleotide sequences from the present invention. AAH55091 to CC AAH55098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention.

N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

Seq Sequence 5024 AA;

Query Match 5.3%; Score 450; DB 22; Length 5024;

Best Local Similarity 20.6%; Pred. No. 5.6e-11;

Matches 360; Conservative 268; Mismatches 653; Indels 466; Gaps 84;

QY 107 NSRRTPNSDSSDAKSYADLKHVRVYLLTIKELKYPOLF--DLTNHMLTLCN---- 160

DB 3011 NKAALNNDMMKKLDIVAQQDNVRGNNY---INEDSTPONWYNDTINHAQSIIDOVAMP 3067

QY 161 -----IHGFYLI---DGYBEINELLYKLFYDLIRAKLNDVCANDYCQIPFNL 207

DB 3068 TMSHDEIENAINIKHAINALDGEHKLQAKENAN-----LLNSLD----- 3110

QY 208 KTRANELDLVKLV--FGYRKPLDNKDNVGMEDYIKKNKTTIENINELIEESKTTIDK 265

DB 3111 -LNAPODAINLVNEAOTREKVAEQLSAQALNDAMKHLRNSIQNSSVRQESKYI--- 3166

QY 266 NKNATKEEKKLYQAQYDLSYLNQLEAHNLISVLEKRIDTFLKKNENIKELDKINEI 325

DB 3167 NASDARKEO-----YNHAVREVENIINEQHPTLD-----KEIKQLTDAVNO- 3208

QY 326 KNPPANSQNTPTLLDKNKKTEE-----HEKETKEIAKTIKEN-----IDSLFTD 371

DB 3209 -----ANNDLNGVELLDADQNAHOSIPTLMLHNAQAQNALNEKINNNAVTRAKVAIIQ 3263

QY 372 PLELEYLREKNKNIDISAKVETKEST-----EPN---EYPNGVTY-----PLS 412

DB 3264 AKILDHAM--ENLEESIKDKQOVKSSNYINEDPDQVQETYNNAVHVTEILNQTNPULTS 3321

QY 413 YNDINNALNELNSF-----GDLINPDYIT--KEPSK-NIYTDNERKKFINET--REKIKI 462

DB 3322 IEDIEHAINEVNAQKQKRGKOKLYQITDLADKLEKSLDDLLTSQOSSISNOLYTAKTAT 3381

QY 463 EKKKTESDKSYEDRSKSLNDITKEYELLNIEYDKFNNNIDLTNFEKMMGKRYSYKVE 522

DB 3382 EVAQAIEKAKSLNHAMKALNKIYKNADKVID---SSRFINE---DOPEKEAQQAINHVD 3435

QY 523 KLTHHT-----FASYENSKHNLEKLTALKYMEDYSLRNI-----V 559

DB 3436 STIHRQTNPEMDPTVINSTHETEAQNNLHGDQKLAHAKQ-----DRAVNLINGLIHLNV 3490

QY 560 VEKEL-----KYKKNL--ISKIENETIETVE-----NIKKDEOLFEEKITK 599

DB 3491 AOREVMINTNTNATPREKAVKALNDNAQALDKAMETLQQVVAHKHNLNDSKYLNEO--SK 3548

QY 600 DENKPEDEKILEVSDIVKVOQKVLNKNKIDELKKTQL-----IL-----KNVEL 643

DB 3549 YQQQYDRIADAEOQLLNTTNPTEPYKVDIVKDNVLANEKILFGAEKLSYDKSNADEI 3608

QY 644 KNIHVPNSYKQENKQ-----EPYLIYLKKEIDK-LKVFMPKVESLINEEKKNIK 693

DB 3609 KHMNLLNNAQKOSIKDMISHAALRTEVKQLQQAOKTLDAMKSLDKTQVWITDTLPNY 3668

QY 694 TEGQSDNSEPSPGETITQATTGPGQAAGSALEGSVOAQAOEQQAQPPV--PVPVPPA 751

DB 3669 TEASEDKKEKVDQTVSHAQAIIT--DKINGSNVSLQOVROALQQLTQASENLQDQVBEA 3726

QY 752 KAAQVTPPAPVNNKNTENSKLDYLEKLYEFLNTSYICHKYILVSHSTMNEKILKQYKITK 811

DB 3727 KV-----HANQITDQLTHLSLQO-----QTAKESV---KNATK 3757

QY 812 EESKLSGCDPLDLLENFNTONNIPVMSWPDLSNLSOLFMEIYKEVMVCNLYKLKDNDK 871

DB 3758 LEEIATASNAL-ALNKVMGKLEQFINHADSIENS-----DNYRQADDDK 3801

QY 872 I---KNLLEAKKV--STSVKTLSSSQMPLSLTPQDKPEVSANDDTSHSTLNLSL-KL 925

DB 3802 ITAYDDALEHGDIOKSNATONEAKAQLQO-----INAETSLNGFERLNHARPPA 3852

QY 926 FENILSLGKNKIYQELIGQKSSNF---YEKILKDSOT-----FYNESEFTNEVSKK 974

DB 3853 LEYIKSLEKINNAQKSALEDKVTQSHDLLEHLVNEGTNLNDIMGELANAIIVNNYATPK 3912

QY 975 ADDINSLNDESKRK-KLEEDINKLKTLL-----OLSFPLYNKYKLKLERLFDKKTGCK 1027

DB 3913 A-SINYINADNRKONFTQAINNARDALNKTOGQNLDFNAIDTFK---DDIFTKDALN- 3967

QY 1028 YKMQIKKLTLLKEQLESKLNSLNNPKHVLQNFVSFFNKKK---EAEIAETENTLENTKI 1083

DB 3968 ---GIERLTAASKAEKLIDSLK-----FINKAQFTHANDEINTNTSIAQLSRI 4013

QY 1084 LLKHY-----KGLVYKYNAGESSPKLT-----SEESIOTEDNYASLENKFKVLKLEGL 1132

DB 4014 VNOAFDLNDAMKSLRDELNNQAFVQASSNYINSDLDLQKQFDHALSNARKVLAKENGKN 4073

QY 1133 KDNLEKKKLSYLSGSLHLLIAELKEV-----IKKNYTCNSP---S 1172

DB 4074 LDEIQIE-----GLKQVIEDTKDALNGTORLSKAKAKAIQVQSUSYINDAQRHIA 4124

QY 1173 ENNTDVNNALESYKKFLPEGTQVATVWSGSDTLEQSQPKKPASTHVAESNTITTSN 1232

DB 4125 ESNINSDLSLANTLSKASDLDNAMKDL-RDTLESN-----STSVPSNVNYINADKN 4177

QY 1233 VDEYDDVI-----IVPIFGESEEDYDDLQGVVTEAVTPSVVIDNLSK 1276

DB 4178 LQIEFDEALQQAASATSSKTSNPATIEEVLGLSQAIY-DTKNALNGE-----OR 4225

QY 1277 IENEYEVLYLKLPLAGYRSLLKQLENNVMTFNVNVKDIILNSRKNENKFNKVESLPIY 1336

Db 4226 LATE-KSKDLKLIKGLKDLNKAQLE-----DVTN-----KVSANTLFE-----L 4264

Qy 1337 KDLTSSNVVVKDPYKFLNKKERDRFLSSVYIKDSI---DITDFINFANDVLGYKILSEKY 1393

Db 4265 SOLQTSTLKLNDKMKLL-----RDLKTLVNPVKASLNVNRADYN-----LKRQF 4309

Qy 1394 KSDLDSTIKKYINDKOGENEKYLPLFNITETLYKTVND-----KIDLFVI 1437

Db 4310 NKALKEAGVGNKNSGTNVN---INDIOHLLTQIDNAKDQDLNGERRLKEHQKSEVFII 4365

Qy 1438 HLEAKVLNVTYE-----KSNVEVKI-----KELN-VLKTIDOKLADF-----KKN 1477

Db 4366 K-EDILNNAOKAALINOIRASKDIKIINOIVDNAIELNDAMQGLKEHVAQLTATTKDNI 4424

Qy 1478 NFVGT-ADLSYDYNH-----NNLLTKFLSTGMVFENLAKTVLSNLLDGN--LOGMLNISO 1529

Db 4425 EYLNADEDLKIQDYAINLANNVLDKENGTKN-DANIIIGMIQNMDDARALLNGIERLKD 4483

Qy 1530 HQCVKQCPQNSGCFRHLDEREECKLLNFKQBGDKCV-ENPNPTCNENNGCDAKCT 1588

Db 4484 AQ-TKAHNDIKDTLKRQDLDETEHANATSNKAAQAKQVMYNEBARKAFSINHATSNDLVNQ 4542

Qy 1589 EEDSGSN 1595

Db 4543 AKDEGQS 4549

RESULT 28

ABP38314

ID ABP38314 standard; Protein; 10182 AA.

XX AC ABP38314;

XX 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.

DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;

KW Staphylococcus epidermidis; gene therapy.

XX Staphylococcus epidermidis.

OS Staphylococcus epidermidis.

XX US6380370-B1.

XX 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR N-PSDB; ABN90859.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis

PT polypeptide, useful for diagnosing and treating bacterial infections -

XX Disclosure; SEQ ID 3159; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading

CC frame (ORF) nucleic acid sequences which encode the amino acid sequences

CC given in ABP35124 to ABP37960. The S. epidermidis sequences have

CC antibacterial activity and can be used in gene therapy. The sequences

CC can also be used in the diagnosis and treatment of bacterial infections,

CC particularly S. epidermidis infections. The sequences can be used to

CC screen for compounds able to interfere with the S. epidermidis life

CC cycle or inhibit S. epidermidis infection.

CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC USPTO web site.

XX Sequence 10182 AA.

Qy Query Match 5.3%; Score 448; DB 23; Length 10182;

XX Best Local Similarity 20.6%; Pred. No. 1.6e-10;

XX Matches 364; Conservative 250; Mismatches 651; Indels 498; Gaps 81;

Qy 107 NSRRTNPSDSSDAKSYADLKHRVRYLLTIKELYPQLF--DLTNHMLTLCN----- 160

Db 7957 NKALANDMKKLKDIVAQODNVROSNNY---INEDSTPONMYNDTINHAQSIIDQVNP 8013

Qy 161 -----IHGPKYLI---DGYEINELLYKLNFYEDLLRAKLVNDVANDYCOIPENL 207

Db 8014 TMSHDEIENANNIKKHAINALDGEHKLQOAKENAN-----LLINSUND----- 8056

Qy 208 KIRANELDVAKLV--FGYRKPLDNKIDNVGKMDYIKKKKKTENINELIEESKTTDK 265

Db 8057 -LNAPORDAIIIRLYNEAQTRKVAEQLOSQAALNDAMKHLRNSIQNSVROESKYI--- 8112

Qy 266 KKNATKEBEKKKLYQAQYDLSIYNKQLEEAHLISVLEKRIIDTLKKENIKELLDKINEI 325

Db 8113 NASDAKKEQ-----YNHAREVENIINEQHTLD---KEIIKQLTDGVNQ- 8154

Qy 326 KNPPANSNGTPTLLDNKKIE-----HEKEIKIAKTIKFNIDSLFTDPL----- 373

Db 8155 -----ANNLNGVELLDADKQNAHQSIPTLMHLNQAQNALNEKINNNAVTRTEVAALIQ 8209

Qy 374 -----ELEYLREKNKNIDISAKVETKESTEPNEYPNGVY-----PLSY 413

Db 8210 AKLLDHAMENLEESIKDK-EQVKOSSNYINEDSDVQETYNADVHVTILNQTVMPTLSI 8268

Qy 414 NDINNALNELNSF-----GDLINPFDT--KEPS-NIYTDNERKKKFINET-KEKIE 463

Db 8269 EDIEHAINEVQAOKQLRGKQKLYQTIDLADKELSKLDDLTSQQSSSSISNIIYAKTTE 8328

Qy 464 KKKIESDKKSYEDRSKSLNDITKEYELKLNIEIYDSKFNNIDLTNFEKMMGRYSYKVEK 523

Db 8329 VAQAIERAKSLNHAMKALNKYKNADKVLDD---SSRFINE---DQPEKKAYQAQINHVDS 8382

Qy 524 LTHNT-----FASYENSKHLEKLTALKKYMEDYSLENI-----VV 560

Db 8383 IIRQTNPMPDPTVINSTHETEAQNNLHGDQLAHA-----QODAAVINGLIHLNVA 8437

Qy 561 EKEL-----KYKNL--ISKTEIETIVE-----NIKKDEEOLFEEKTKD 600

Db 8438 QREVMINTNTATTREKVAKNLDNAQALDKAMETLQOVVAHKNNILNDSKYLNE--SKY 8495

Qy 601 ENKPDEKILEVSDIVKVOQVQLMLNKKIDELKKTQL-----IL-----KNVELK 644

Db 8496 QQQYDRVIADAEQLLNTTNPTEPYKVDIVKDNVLANEKILFGAEKLSYDKSNANDEIK 8555

Qy 645 HNIHVPNSYKQENKO-----EPYLLIVLKKKIDK-LKVFMPKVESLINEEKKNTKT 694

Db 8556 HMNYLNNAQKOSIKDMISHAALRTEVKOLQOAKILDEAMKSLEDKTQVITDTPLPNT 8615

Qy 695 EGQSDNSEPSTEGETTGQATTKPGQAGSALEGSVQAQAEQQAQPPV--PVPVPEAK 752

Db 8616 EASEDKKEKVDQTVSHAQAI--DKINGSNVSLDQVRAQLQEQTOASENLGDQORVEAK 8673

Qy 753 AQVPTTPAPVNVKNTENYSKLDYLEKLYEFLNTSYICHKYILVSHSTMKEKILKQYKTYKE 812

Db 8674 V-----HANOTIDOLTHLSL-----QOOTAKE 8696

Qy 813 EESKLSCCDPLDLLENIONNIPVMYSMFDLSNLSQLFMEIYKEMVCNL--YKLDND 870

Db 8697 ---SVKNATKLEETATVSNN-----QALNKVNGKLEQFINHADSVNSDNYRQADDD 8746

Qy 871 KI-----KNLEAKKYSTSV---KTLSSSSMOPL-----SLT 900

Db 8747 KIIAYDEALEHGDIQKTNATQNETKQALQQLIYAETSLNGFERLNHARPRALEYKSL 8806

Qy 901 PQDKPEVSANDTSHSTNLNLSKLFINLSLGNKNKIYQELIG-----KSS 948

Db	8807	KINNAKSALED---KVTQSHDLLEHIVNEGTLN---DIMGELAINVINYAPTRAS	8860	
Qy	949	ENFYEKILKSDTFYNESFTNFVSKADJNSLN---DESKRKKLEEDINKLKTLOLS	1004	
Db	8861	INYI-----NADNLKDNFTQAINARDALNKTQGNLDFNADTFKDDIFKTDALN-G	8914	
Qy	1005	FOLYNYKILKLERLFDKKTGVGYKW-----QIKKLT-----LKEQLESKL	1046	
Db	8915	IERLTAASKABKLDLSLKFINKAQFTHANDEIMNTNSTIAQLSRIVNQAFDLNDAMKSLR	8974	
Qy	1047	NSLNNPKHVLQNFVSFFNKKKEAETAENTLENT-KILLKHVGLVYKYGESSPLKTL	1105	
Db	8975	DELNNQAFVQASSNYINSDEDLK-QQFDHALSNARKVLAK-----NG-----KML	9020	
Qy	1106	SESTQ-----TEDNYASILENFVKLSLEGLKDNLNLEKKLSYLSGLHHLIAELREV	1160	
Db	9021	DEKIQGLQKVETDKDALGQIRLSKAKAKAIQV-----QSLSYINDAQRHI-----	9069	
Qy	1161	IKNKYNTGNPSSENNTDNNALSKYKFLPECTGVATVVSSESGSDTLQSQPKPASTHV	1220	
Db	9070	-----AENITHSDOLSLANTLSKASDLDNAMKOL-RTIESN-----STSV	9111	
Qy	1221	GAESNTITTSQNVDEVDVI-----IVPGESEEDYDGLGQVVTGEA	1264	
Db	9112	PNSVNYINADKNLQIEFDEALQOASATSKTSENPATIEEVGLSQAIY-DTKNALNGE-	9169	
Qy	1265	VTPSVIDNLSKIENEYEVLYLKLPLAGVYRSLLKOLENNVMTFNVNVDILNSRFNKREN	1324	
Db	9170	-----QRLATE-KSKDLKLKGLKDLNKAOLE-----DVTN-----KVN	9202	
Qy	1325	FRNVLESLIPYKDLTSSNYVVKDYKFLNKKRDKRFLSSNYIKDSI-----DTDINFAND	1381	
Db	9203	SANTLTE-----LSQLTQSTLELNDKMKLL-----RDKLTLNVPKASLNRYNADYN----	9250	
Qy	1382	VLGYKILSEKYSKLDLSIKKINDKOGENEKYLPLFNNIEFLYKTVND-----	1430	
Db	9251	-----LKROFNKALKEAGVNLKNKSGTNV-----INDIHLTLQIDNAKDOLNGERRL	9299	
Qy	1431	-----KIDFLVHLBAKLVNYE-----KSNVEKI-----KELN-----YLKTI	1466	
Db	9300	KEHQKSEVFIK-ELDLILNNAKAAIINQIRASKDIKIINQIVDNAIELNDAMQGLREH	9358	
Qy	1467	QDKLADFKNNFVGIADLSTDYNH-----NNLLTKFLSTGMVFENLAKTVLSN	1515	
Db	9359	VAQLTATTKDN-----IEYLNADHDHKLQDYAINLANNVLDKENGNTK-DANIILIGMIQN	9413	
Qy	1516	LLDGN--LOGMLNISQHCVKQCPQNSCCFPHLDERECKLLNYKQGGKCV-ENPNP	1572	
Db	9414	MDARALLNGIERLKDAQ-TRAHNDIKDTLKRQLDDEIEHANATSNKAAQAKQMVNEEAR	9472	
Qy	1573	TCNENNGCDADAKCTEEDSGSN	1595	
Db	9473	ALSINDATSNLDVNOAKDEGQS	9495	
RESULT 29				
ID	ABG06505	standard; Protein; 2633 AA.		
XX	AC	ABG06505;		
XX	DT	13-FEB-2002 (first entry)		
XX	DE	Novel human diagnostic protein #6496.		
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
XX	KW	food supplement; medical imaging; diagnostic; genetic disorder.		
XX	OS	Homo sapiens.		
XX	PN	WO200175067-A2.		
XX	XX			

PD	11-OCT-2001.			
XX	30-MAR-2001; 2001WO-US08631.			
XX	31-MAR-2000; 2000US-0540217.			
PR	23-AUG-2000; 2000US-0649167.			
XX	(HYSE-) HYSEQ INC.			
PI	Drmanac RT, Liu C, Tang YT;			
XX	WPI; 2001-639362/73.			
DR	N-ESDB; AAS70692.			
XX	New isolated polynucleotide and encoded polypeptides, useful in			
PT	diagnostics, forensics, gene mapping, identification of mutations			
PT	responsible for genetic disorders or other traits and to assess			
PT	biodiversity			
XX	Claim 20; SEQ ID No 36864; 103pp; English.			
PS	The invention relates to isolated polynucleotide (I) and			
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,			
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome			
CC	and gene mapping, and in recombinant production of (II). The			
CC	polynucleotides are also used in diagnostics as expressed sequence tags			
CC	for identifying expressed genes. (I) is useful in gene therapy techniques			
CC	to restore normal activity of (II) or to treat disease states involving			
CC	(II). (II) is useful for generating antibodies against it, detecting or			
CC	quantitating a polypeptide in tissue, as molecular weight markers and as			
CC	a food supplement. (II) and its binding partners are useful in medical			
CC	imaging of sites expressing (II). (I) and (II) are useful for treating			
CC	disorders involving aberrant protein expression or biological activity.			
CC	The polypeptide and polynucleotide sequences have applications in			
CC	diagnostics, forensics, gene mapping, identification of mutations			
CC	responsible for genetic disorders or other traits to assess biodiversity			
CC	and to produce other types of data and products dependent on DNA and			
CC	amino acid sequences. ABG00010-ABG30377 represent novel human			
CC	diagnostic amino acid sequences of the invention.			
CC	Note: The sequence data for this patent did not appear in the printed			
CC	specification, but was obtained in electronic format directly from WIPO			
CC	at ftp.wipo.int/pub/published_pct_sequences.			
XX	Sequence 2633 AA;			
Query Match				
Best Local Similarity 5.3%; Score 445; DB 22; Length 2633;				
Matches 334; Conservative 327; Mismatches 610; Indels 486; Gaps 78;				
Qy	126	ADLKHVRVNYLTITKELK---YPQLFDLTNHLMT-----LCDNTHGPKY	166	
Db	415	AKRRRVTWCLGKINKMKNNSYADQFNPTNTTTTKHLSINLLREIDESVCSDFSN	474	
Qy	167	LIDGYEEI-----NELLYKLNFFDL--LRKLNDCANDYCOI-----PFNLKTR-ANE	213	
Db	475	TLDTLSEIENWPATKLLNQENIESSELNSLRADYNLV-LDYQLRTEKEEMELKURKND	533	
Qy	214	LDV-----LKLIVFG---YRKPLDN-----IKDNVKG	237	
Db	534	LDEFALERKTKKQEMQLIHEISNLKLVKHREYVYNQDLENELSKSVKVELLREKEDQIKK	593	
Qy	238	MEDYIKKKNKTTIENIN-----ELIESKK-----TID-KKNKATKEEKKKL	278	
Db	594	LOEYIDSQK--LENIKMDLSYSLESTEDPKQMKTLDFAETVALDAKRESAFRLSENLEL	651	
Qy	279	YQAQYDLSYNNKQLEAHNLISVLEKRIDTLAKN--ENIKELDKDKNETKNPPANSNGT	336	
Db	652	KEKMKELATTYKOME---NDIOLYOSQLEAKKKMQVDLEKELOSAFNETKITLTSIDGKV	708	
Qy	337	PNTLL-----DNKKITEEHF-----KEKEIARTIK----	362	
Db	709	PKDLICNLNLECKITDLOKELNKEVEENALREEVILLSELKSLPSEVERLKRKEIQDKSE	768	

Qy 363 -----FNIDSLFDPLELEY-----YLREKNKNIDISAKVETKESTPEPNEYPNGVITYPLS 412
Db 769 ELHIITSEKDKLPSEVVKESRVOGLLEIGTKTKDLDLATQSNYSKSTQDFQNFKTLHMD 828
Qy 413 YN-----DINNALNEL-----NSFGDLINPDY-TKEPSKNIYTDNERKKF 452
Db 829 FEQYKVMLEENRMRNGEIVNLSKEAQKFDSSGALKATSELSYKTOELQEKTRVQERLNE 888
Qy 453 INEIKKIK-----IEKKK-----IESDKKSYEDRSKSLNDITKEVEKL 491
Db 889 MEQKQELNDRSPLOVVEREKLITTEKLOOTLEEVATLTOEKDDLKQLOESLOIERDQ 948
Qy 492 LNEIYDSKFNNDITNFKEMKGRYSYKVEKLTHH-----NTFAS-----YEN 535
Db 949 KSDIHDY-VNMNIDTQ-----EOLRNALLESKQHOQETINTLKSISSEVSRNLHMEEN 1000
Qy 536 ---SKHLEKLTALKWEDYSLRNI-----VWEKELYKYNLISKIENETJL 581
Db 1001 TGETKDFQOKWGDIDKKQDLKAKNTQTLTADVKDNEIIEQQRKIF-SLIQE-KNELQOM 1058
Qy 582 VENIKKDEEOLFEEKITKDN-----APDEKILEVSDIVKVOQVLLMKNKIDELKKQTOL 636
Db 1059 LESVIAEKEQL---KTOLKENIEMTINQELRLIGDELKKQ-OEIVAQEKNAHAKKEGE 1114
Qy 637 ILKN---VELKHNHVPNSYKQENKQEPYLIIVKKKIDKLVFMPKVESLINEER--- 689
Db 1115 LSRCTDLAEVEEKLKEKSQOLQEQKQ---LLNVQEMSEMOKKINEIENLKNELANKE 1171
Qy 690 ---KNIKTE---QSONSEPSPEGETGQATPKPGQOAGSALBGD-----SVQAOAQ 735
Db 1172 LTLHEMETERLELAQKLNENYEEVKSITKERRKVLKELQKSFETERDHLRGYIRIEATGL 1231
Qy 736 EQQAOQPPVPVPEAKAQPPTPAPVNNKTNVSKLDYLEKLYEFLNTSYICHKYLVS 795
Db 1232 QTRKELIAHILHEKHETIDELRSVSEKTAQIINTQDLEK-----S 1274
Qy 796 HSTMNEKILKOYKITKEESKLSKCDPLDLFINQNNIPVMSFDSLNNS----- 846
Db 1275 HTKLQEEI---PVLHEEQELLPNVKVSQETQETMNELELLITQSTTKDSTTLARIEMER 1330
Qy 847 --LSOLFNEIYEKEMVNLKLDN-----DKKNLLEAKKYSTSVKTLSSS 892
Db 1331 LRUNEKFOE---SQBEIKSLTKERDNLTKIEALEVKHQDLKEHIRE-----TLAKIQESQ 1383
Qy 893 SMOPLSTPDQKPEVSANDTSHSTNLNNSKLPE-----NILSLGKNKNIYQELI 943
Db 1384 SKQEOSLNMKEK-----DNETHKIVSEMEQPKPDSALLRIEIMGLSKRL----- 1430
Qy 944 GOKSSENFYKILKDSFTFYN-----ESFTNFVKSADDINSLNDESKR-----KK 989
Db 1431 --QESHDEMSVAKKDDQLRLQELVQESDQLENKEIVAKHLETEELKVAHCCLKE 1488
Qy 990 LEEDINKLK-----KTLQSLFDLYN-KYKLLERLFDKK-----KTVGKYKMQIK 1033
Db 1489 OEETINELRVNLSEKETEISTIQKLEAINDKQNKQTOEIEKEEQNLKQISEVQENVN 1548
Qy 1034 KLTLLKE-----QLESKLSNLSNPKHVLFQNFVSFFENKKK-----EAEIAETEN 1076
Db 1549 ELKQFKEHRAKDSALQIESKMLELTNRLOESQEEIQLMIKEEMKRVQEAQIERDQ 1608
Qy 1077 TLENTKILLKHGVLKYNGESPLKLTSEESI-QTEDNYASLENPKVLSKLEGLKDN 1135
Db 1609 LKENTKEI-----VAKMKESQEKEYQFLKMTAVNETQEMCEIEHLKEQFETQKLNLEN 1662
Qy 1136 LNLKLLSYLSLGLHLLIAELKEVINKNYTGNPSNNTDNNALLESYKKFLPEGTDV 1195
Db 1663 IETENIRLTQI---LHENLEEMRSVTKERD-----DLRSVEETLK---VERDQL 1705
Qy 1196 ATVVSSEGSOTLEQSPKPPASTHVAESNTITTSQNVDDVEDDVIIVPIFGESEEDYDD 1255
Db 1706 KENLRITRDLQKQELKIVHMLKHEQETIDKLRGIVSEKTNEL-----SNMOKOLEH 1760
Qy 1256 LGQVVTGEAVTPSIDNLSKIENEYEVLYLKPLAGVYRSLKOLENNVMTFNVNVKDIL 1315

Db 1761 SNDALKAQDL-----KIQBELRIAHM-----HLKEQOE-----TID-KLRGIV 1797
Qy 1316 NSRENKRENKAVNLESLLIPYKDLTSSNYVVDYKPKFLANKERKDFLSSYNYIKDSIDTD 1375
Db 1798 SEKTDKLSNMQ-----KOLENSNAKLOEKIOELKANEHQLI-----TLKKD 1838
Qy 1376 INFANDVLGVYKILSEKYKSDLSIKKYINDK-----QGENEKYLPFLN-NIETLYKT 1427
Db 1839 VN-----ETQKVVSEMEQLKKQIKDOSLTSKLEIENLNAQELHENLEEMKSV 1887
Qy 1428 VNDKIDILFVHLEAKVLNYTEKSNVEVKIKELNY-----LKTIOQDLADFKNNNFVGAD 1484
Db 1888 MKERDNLRRVEETLKERDQLESQETKARDLEITQOELKTCSEKISE-----KTIQISD 1942
Qy 1485 LSTDY--NHNLLTKFLSTGMVFENLAKTVLSNLL-----DGNLOGLMNLISOHCYKQCPQ 1539
Db 1943 IQDLDSKDELOKK-----IQELOKKEQLLRVKEVDVNMHSK-KINEMEQLKKOPEP 1994
Qy 1540 NSGCFRHLDERECKCL 1556
Db 1995 NYLCKCEMDFOLTKKL 2011
RESULT 30
AAM40883
ID AAM40883 standard; Protein; 2688 AA.
XX AAM40883;
AC AAM40883;
XX 22-OCT-2001 (first entry)
DT Human polypeptide SEQ ID NO 5814.
DE
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
OS WO200153312-A1.
XX 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US34263.
PF 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI60039.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Example 2; SEQ ID NO 5814; 10078pp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

the encoded polypeptides (AAM38642-AM42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

Sequence	2688 AA;
SQ	

Query Match 5.38; Score 444.5; DB 22; Length 2688;

Best Local Similarity 18.4%; Pred. No. 4.6e-11;

Matches 345; Conservative 350; Mismatches 616; Indels 559; Gaps 82;

Qy	126	ADLKHVRNVLITIELK----	YPQLFDLTNHMLT-----	LCONIHGFKY	166
Db	439	AKBKRVTWCLGINKKNMSYADOFNIPTWTTTKHLSINLLREIDESYCSESDVFN	498		
Qy	167	LIDGYEEI-----NELLYKNFYDPL--LRAKLDNVCAODYCOT-----PENLKIR-ANE	213		
Db	499	TLOTLSIEWNPATKLNGENTISELSNRADYDWLV-LDYEQURTEKEEMELKAEKEND	557		
Qy	214	LDV-----LKLTVFG---YRKPLDN-----IKDNVGK	237		
Db	558	LDEFEALERTKKDOEWOLHEISNLKNLVKHREYVNOGLENSKVVELLREKEDQIKK	617		
Qy	238	MEDIYKNNKTENIN-----ELIESKK-----TID-KNKNAETREEKKKL	278		
Db	618	LOEYIDSOK--LENIKMDLSYESIEDPKOMQTLPDAETVALDACRAFLRSENLEL	675		
Qy	279	YQAQYDLSTYNKOLEAHNLISVLSEKRDDTLKKN--ENIKELLDKINEIKNPPPANSNT	336		
Db	676	KEMKELATTYKOME---NDIOLYOSQEANKMQVDLEKELQSFAFNEIKTSLIDCKV	732		
Qy	337	PNTLL-----DNKKKTIEEH-----KEIKEITARIK---	362		
Db	733	PKDLLCNLEGGKITDLOKELNKEVEENEALREEVILLSELKSLPSEVERLKEIQDKSE	792		
Qy	363	-----FNDSLUFTDPLEXY----YLREKKNNIDISAKVTKESTEPNPGVTPVPLS	412		
Db	793	ELHIITSEKDCLKLFSEVVHKRESRVQGLLEEIGTKTDLATTSKSTQDOEFQNFKTLHMD	852		
Qy	413	YN-----DINNANLEL-----NSFGDLINPFDF-YKEPSKNIIYTDNERKFF	452		
Db	853	FEOKYKMVLSEENERMGOEIVNISKAQFPDSLGAUKTELSTYKTOELOEKTREVQOEKNE	912		
Qy	453	INIEIKKIK-----IEKKK-----IESDKSYESDRKSJLMDITKEYEKL	491		
Db	913	MEQLKQLENRSPLQTVBREXKTLITEKLOQTLEBVKLTQEKDDLQLOESQIERDQL	972		
Qy	492	LNBIYSOKFNNDITLNFEMMGKRYSKYEKLTHH-----NTFAS-----YEN	535		
Db	973	KSDITHDT-VNMNIDTQ-----EQLRNALESJKOHOETINTLASKTISEEVSRLNHHNE	1024		
Qy	536	--SKHNLEKLTKALKYMEDYSIRNI-----VVEKELYKNKLISKIBNETIETL	581		
Db	1025	TGETKDEFQKWGVGDKKODLEAKNTQTLTADVKNONEIEQORIKIF-SLIQE-KNEUQQM	1082		
Qy	582	VENIKKDEOLEFKKTIKEN-----KPDEKILEYSDIVKVQOVKVLMMKNIDBLKCTLQ	636		
Db	1083	LESVIAEKEQL---KTDLENIENTIENOEBELRLIGDELKKQ-QEIVAQEKHAHAIKKEGE	1138		
Qy	637	ILAN-----VELKHNIHVPNYSKOENKQBYYULIVLUKKEIDLKLVMPKPVKESLINEEK	--- 689		
Db	1139	LSRTCORLAEBEKLKAEKSOQLOERKOOO---LNNYOEMSOMOKKINIEINLKNLKNKE	1195		

QY 966 SFTNEVKS-----KADDI-----NSLND-----ESRRKKLE 991
DB 1271 HLPNEKSYNIVYVEDISFDIFLISIMDLWETNNNNLLNLINDLLKIYEERKKKIY 1330
QY 992 -----EDINKKTLQSLFDLYNKKYKLERLFD-----KKTVGVKYK----- 1029
DB 1331 ICTSLLLKIFRIKKKSNYFLNIFKAFENDIKLILDSINILIKKVVWTFKNCNIF 1390
QY 1030 -----MQIKKTLLEQELSKLNSLNPKHVLONSFVFNKKKEAIEATENTLENTK-- 1082
DB 1391 NREKNINIKK--LVKLFIFISFYKYLKN--YFQIYYHFF-----YNNQIYNRKNY 1436
QY 1083 -----ILLKHY--KGLVKVYNGESSPLKTLSESIOTEDNYASLENFKVLSKLEGLK 1133
DB 1437 NFDNFFSFSKYINKIFVEIYSSSSSTSSNSFVFNKSFYMMKMCISIINNMGVVK 1496
QY 1134 DNLNLEK--KLSY-----LSSGLHLHIAELKEVIKKNKYTGNSPSENNTDNNALE 1183
DB 1497 -YINLERVKQFYEHNNIMDMVHMKSHLHDI-----DVYGHDSYNNI-YQKIHK 1545
QY 1184 SYKKFLPECTVATVVSSEGSTLEQSPKPKPASTHVAESNTITTSQNVDDEVDDVIV 1243
DB 1546 SYR---GEEKDLDVNI-----TESVHONRNEDDIDGSINS 1579
QY 1244 PIFGESEEDYDLGQWVTGEAVTPSVIDNILSKIENEVEVLYKPLAGVYR---SLKKQ 1299
DB 1580 DVFNE-----IMRNIIINHNSLIKDHNDMCTKKKRINFQISSPATSEQ 1623
QY 1300 LENNYMTFNVNVKDIILNSRFNKNFNKVNLESDLIPYKDLTSSNVVYKDPYKFLNKKERD 1359
DB 1624 LMNNHTWNY-LTDVW-----LLQKDYIYNID-----NNMNEHKQN 1658
QY 1360 KFLSSYNIKOSIDIDFANDVLCYKILSE---KYKSDLDLSIKKYINDKQGENEYKL 1415
DB 1659 VFNKPF-----DNNNNNNNNNNMFLNFFYPENNNNNYRMD---IKKRYPPESYDNNYIM 1711
QY 1416 PPLNNIETLYKTVNDKIDLV-----IHLEAKVLYNTYKESVNEVKIKEL-NYLKTI 1466
DB 1712 ---FNNIK-----NEENILLQNNSSSIYIDKLMKDTKEMEPLFNKTKDMKNY--- 1759
QY 1467 QDKLADFKNNFVGIADLSTYDINNLLTK-----FLST-----GMVFENLAKTVLSN 1515
DB 1760 ----NEEQNNELI-----SYPNMNLQNNIIFVKFFLYTQNLQIIFQNNYIFFLSD 1808
QY 1516 LLDGNLQGLMLNISOHCV--KKOCFQNSGCFRHLDEREECKLLNYKOGDKCVENPNPT 1573
DB 1809 FL-----FINKKYEIEEKNQNV-----INIKDEKHTNIK-DGDKHITNIKDG 1856
QY 1574 CNEENGGCDADAKTEEDSGSGNKKITC-----ECTKPDSPYPLFDGIFCSSSNFL 1623
DB 1857 DKNITNKDDDKNITNMKKKNKNTILMYNSQECs--FYYSIFNTLI-NDYNFL 1909
RESULT 32
AAM39097
ID AAM39097 standard; Protein; 2663 AA.
XX
AC AAM39097;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2242.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
OS Homo sapiens.
XX

PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AAI58253.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 2242; 10078pp; English.
PS
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2663 AA;
Query Match 5.2%; Score 441; DB 22; Length 2663;
Best Local Similarity 18.8%; Pred. NO. 6.5e-11;
Matches 323; Conservative 325; Mismatches 578; Indels 494; Gaps 74;
QY 126 ADLKURVNYLLTIKELK--YPQLFDLTNHLMT-----LCDNIHGFKY 166
DB 415 AKRRRVTCWLGKINKMNSNADQFNPTNTTTTKLSINLLREIDSVCSSESVFSN 474
QY 167 LIDGYEEI-----NELLYKLNFDL--LRAKINDVCANDYQCI-----PFNLKIR-ANE 213
DB 475 TLDTLSEIENWPATKLLNQENIESELSNRADYDNLVL-DYEQLRTEKKMKELKEKND 533
QY 214 LDV-----LKLIVFG---YRPLDN-----IKDNVCK 237
DB 534 LDEFEALERKTKKQDMQLIHEISNLKLVKHKREYVYNQDLENLSKSVLLEKEDQIKK 593
QY 238 MEDYIKKKNKTTENTN-----ELIEESKK-----TID-KKNATYKEEKKKL 278
DB 594 LOEYIDSKR--LENIKMDSYSLESTEDPKMQKTLFDRAETVALDAKRESAFILSENLEL 651
QY 279 YQAOYDLSLYNKKQLEEAHNLISVLEKRRIDTLKKN--ENIKELDKINEIKNPPPSNGST 336
DB 652 KEKMKELATYKQME---NDIQLYQSLQEAKKMKQVLDLEKELQSAFNEITKLTSLDGVK 708
QY 337 PNTLL-----DNKKKTEEHE-----KEIKELATIK----- 362

Db 709 PKDLCLNLEGGKTTDLOKELNKEVEENEALREEVILLSELKSPSEVERLKEIQDSE 768
QY 363 -----FNIDSLFDPLEY-----YIREKNKNDISAKVETKESTEPNPGVTVPLS 412
Db 769 ELHIITSEKDKLSEVVHKSRSVQGLEEGTKYKDDLATQSNYSKSDQEFQFKTLHMD 828
QY 413 YN-----DINNALNEL-----NSFGDLNPFYD-TKESKNYITDNERKKF 452
Db 829 FEQYKMLVEENRANQEIYNLSKEAQKFDSSLGALCATSELSYQELQEKTRVQOERLNE 888
QY 453 INETKEKIK-----IEKKK-----IESDKKSVEDRSKSLNDITKYEKL 491
Db 889 MEQKEQLENRDSPLQTVEREKTLITBKLQOTLEEVKTLTOEKDDLKQLESQIERDQL 948
QY 492 LNEIYDSKFNNDITLTFERKMGKRYKYVEKLTTH-----NTPAS-----YEN 535
Db 949 KSDIHDY-VNMNIDTQ-----EOLRNALSLKQHOETINTLSKTSISEEVSRLNLMHEEN 1000
QY 536 ---SKHNEKLTALKYMEDYSURNI-----VVEKEUKYKNLISKIENEIETL 581
Db 1001 TGETKDFQOKMGVIDKDKQLEAKNTOTLTADVKNDBEIEQQRKIF-SLIQE-KNELQQM 1058
QY 582 VENIKDEEOLFEEKITKDEN-----KPDEKILEVSDIVKVOVKVLLMKNKIDELKKTQL 636
Db 1059 LESVIAEKEQL---KIDLKENIEMTIENQBELRLGDELKKQ-OEIVAQEKNHAIKKEGE 1114
QY 637 ILKN---VELKNIHVPNSYKQENQEPYILVLAKEIDKLVKVPKVESLINEEK--- 689
Db 1115 LSRTCORLAEVEERKEKESQOLQEQOO---LLNQVEEMSEMQKKINEIENLAKNELNKE 1171
QY 690 ---KNKTE---GQSDNSPSTGEITGQATKPGQAGSALGEGD-----SVQAAQ 735
Db 1172 LTHEMETERLELAQNLNENYEVKSTKERRKVLKELQKSFETERDHLRGYIREIATGL 1231
QY 736 EQQAQPPVPVPEAKAQPPTPAPVNNKTENVSKLDYLEKLYEFLNTSYICHKYLVS 795
Db 1232 QTEELKIAHILKEHQETIDELRSVSEKTAQIINTQDLEK-----S 1274
QY 796 HSTMNEKILKOYKITREESKSLSCDPLDLFIQNNIPVYMSFDSLNS-----KK 846
Db 1275 HTKLQEEI---PVLHEEQELLPVKKVSETQETMNELELLTQESTTKDSTTLARIEMER 1330
QY 847 --LSQLFMEIYEMVKNLYKLDN-----DKTKNLEBAKKVSTSVKTLSSS 892
Db 1331 LRLNERFOE--SOEIKSLTKERDNLKTIKEALEVKHDQKEHIRE-----TLAKIQESQ 1383
QY 893 SMQPLSLTPQDKPEVSANDDTSHSTNLNSKLFE-----NILSLGKNKIYQELI 943
Db 1384 SKOEQSLNMKEK-----DNETTKIVSEMQFKPKDSALLRIETIEMGLSKRL----- 1430
QY 944 GQKSSNFYVEKILKDSPTFYN-----ESFTNFVKSKADDINSLNDESKR-----KK 989
Db 1431 --QESHDEKMSVAKEDKDLQRLQEVLOSQSDQLKENIKEIVAKHLETEBELKVAHCLKE 1488
QY 990 LEEDINKLK-----KTLQSLFDLYN-KYKLLERLFDKK-----KTVGKYRMQIK 1033
Db 1489 QETINELRVNLSEKETEISTIQKLEAINDKLQNKIQEIEYEKEEQINIKQISEVQEKVN 1548
QY 1034 KLTLKE-----QLESKLSNLNPKHVLQNFVSFNNKK-----EAEIATEN 1076
Db 1549 ELQOFKEHRKAKDSALQIESKMLLELTNRLQESQEEIQIIMIKEEMKRMVQEAQIERTDQ 1608
QY 1077 TLENTKILLKHKGVLKYNGESSPLKTLSEESI-OTEDNYASLENPKVLKLEGLKDN 1135
Db 1609 LAENTEI-----VAKMKSQEKYQFLKMTAVNETQEMCMCEIHLKQFQTKUNLEN 1662
QY 1136 LNLKXKLSYLSGLHLLIAELKEVIKKNKYTGNSPSENNTDNNALSKYKFLPPECTDV 1195
Db 1663 IETENIRLQI---LHENLEEMRSVTKERD-----DLRSVEETLK---VERDQL 1705
QY 1196 ATVVSSEGSDDLQSQPKKPASTH-----VGAESNTITTSQNVDEVDV- 1240

Db 1706 KENLRITITRDEKQOEELKIVHMLHKEHQETIDKLRGIVSEKTNISNMQKDLHSNDAL 1765
QY 1241 -----IIVPIFGESEEDYDDLGOVVTGEAVTPS-----VIDNILSKIE----- 1278
Db 1766 KAODLKTOEELRTAHMLHKEQOETIDKLRGIVSEKTDKLSNMQKDLNSNAKIQEKIQEL 1825
QY 1279 --NEYEVLYLK-----PLAGVYRSLLKQLENNVMTF-----NVNVKDIILN----- 1316
Db 1826 KANEHQILITLKDVNETQKKVSEMEQLKQIKQDQSLTSLKLEIENLNAQELHENLEEMK 1885
QY 1317 SRNKNRNFNVLESDLIPYKDLTSSVYVNDPKYFLNKEKRDYFLSSYNYIKSDITDI 1376
Db 1886 SVMKRDNLRLRVEET-----LKLERDQLKES---LQETKARDL 1920
QY 1377 NFANDVLGYKILSEKYKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKIDLFV 1436
Db 1921 EIQOE-LKTARMLSKHEKTEVDKLRKISEKTIQ-----ISDIQDLKSKDELQKKI 1972
QY 1437 IHLEAKVLYNTYKESNVYEVKIKELNYLKTITODKLADEPKKN 1476
Db 1973 QELQKKEQLLRVKEDVNMSHKKINEMEQLKKQ---PEPN 2009

RESULT 33

AAB18294

ID AAB18294 standard; Protein; 980 AA.

XX AAB18294;

DT 07-NOV-2000 (first entry)

XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:152.

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

XX W0200025728-A2.

XX 11-MAY-2000.

XX 05-NOV-1999; 99WO-US26796.

XX 05-NOV-1998; 98US-0107131.

XX (HOFF/) HOFFMAN S.

XX (CARU/) CARUCCI D.

XX (GARD/) GARDNER M.

XX (VENT/) VENTER J C.

XX Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection.

XX Disclosure; Page 362-365; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera of a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rfinos or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the plasmodium chromosome 2 and the

CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasite lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AA70078 to AA70287 and AA81844 to AA81852 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

XX Sequence 980 AA;

Query Match 5.1%; Score 431; DB 21; Length 980;
Best Local Similarity 23.1%; Pred. No. 5.4e-11;
Matches 263; Conservative 177; Mismatches 343; Indels 354; Gaps 52;

QY 176 ELLYKLNFFDLRAKLVNDVANDYQIPFNLIKIRANELDVLKLVFGYRPLDN----I 231
DB 7 ELILKKNY--EELRLKLNHL-----ECVERDSVKIN--SEKGEKVIYELKELDNDEKII 60

QY 232 KD-----NVGKMEDYIKKNKKTENEL---TEESKKTIDKN---KNATKEEKKKL 278
DB 61 NDLKKNSYQVYKMDYKERNLNLINEINKLFLTEENKMTVERGEMNNKKLEEMKOKN 120

QY 279 YOAQYDLSTYNNKQL-----EEAHNLSVLE-----KRI----- 306
DB 121 KELINLNLDISDELKNCIEQVNSVRNMANVEKKEINIINELQILRMKNMTMRKRISFV 180

QY 307 -----DTLKKENIKELDKINEIKNPPANSGNTPNTLLDKNKKIEHE 351
DB 181 EOEKVLKFKLYTLNNDIFSKEKNDKMDKLVN--NEYKNIIVECLNNYKTEHK--EQIE 237

QY 352 KEIKIATKIFNIDSLETPLELYYREK--NKNIDISAKVETKES--TEPNYEPNGV 407
DB 238 KKIERI--NLTQNYYY-----LKKEYDLKNELEKIEHKKLEHLSHCYEENQ----- 286

QY 408 TPLSYNDINNALNELSGDLINFDYTKPESKNIYTDNERKKFTINEIK----- 457
DB 287 -----KLNEBIKRRNSF--IKNKDRKIDLLTNIENELLKKKEINNIKLMKQNVIKN 336

QY 458 -----EKIE-----KKKIESDKSYEDRSKSLNDITKE 487
DB 337 NQOLKIDKIDENEKNEHVKNLQNELIKRELONKICSIDIECKKEKEDKLNLEDDLLE 396

QY 488 YEKLLEIYD-----SKFNNNIDITNPEKMMGRYSYKVEKLTHHTFASYENS-----K 537
DB 397 KKKCIENLKDELINIKKKMEDKMTNEMDLL-----SNKVELNRIN--KTYEKNIVELN 450

QY 538 HNLEKLTALK-----YMEDYSLRNIVWEKELKYKNLSKISKEIETETLVN-----IKKD 588
DB 451 NELDIVIKKLNDEEFLKEEKKKNIDMVKYKIEYIQIKERENEIDSLKKNQNLHVLKN 510

QY 589 EOLPEKKTITDENKPKILFVSDIV-----KVQVQVLLMNTI----- 628
DB 511 EE-LNEKEIL--KNKYDK--EINMIISQYNNKIOEEKMDLNNKIKSMDOQTHKNQIEMQ 565

QY 629 DELKKTQLIKLVN-----ELKHNIHPVNSKQENKQBPYLIYLKKEIDKLVFMPKVESLI 685
DB 566 EENKELRLKLVNCDMNLQSLIKENEKHMQEKVEEYKKNLLKQDKQDELKNIQYDERI 625

QY 686 NEEKNKITEGOSDNSEPSFTEGTTGQATTPGQAGSALGEDSVQAAQKQKQAPVPV 745
DB 626 ETQNKEME-----DIVNDECEKLKQAK----- 647

QY 746 VVPPEAKAQVTPPPAPVNN--KTENVSKLDYLEKLYEFLNTSYICHKYLIVSHSTMNEKI 803
DB 648 ---INNKLTATNANNMNNMLMDENLKEKD--KKI-----NDLKMOME 686

QY 804 LKQYKITK--EESKLSLSCDPLDLFLFNQNNIPNMYSMFDSLNNLSQLFMEIYKEMVC 861
DB 687 KKKEEINKLVEKSKLEHSH-----VKIQNEMSLLVQNEKL-----KEEWGL 729

QY 862 NLYLKDNKDKNLL---EAKKVKSTSVKTLSSSSMQPLSLTPQDKPEVSANDTSHSTN 918

DB 730 SRIATKDMEEIKDKMEKYEKKKKNEERKKNEERK-----KNEERKKNEE----- 778

QY 919 LNNSLKLEENILSLOKKNINIQELIGOKSSNFYKILKDSDTFFYNESFTNFVKSKADDI 978

DB 779 -----KKNEEE-----RKNEEKKKKLEKDKHQFEE-----KERMEIY 813

QY 979 NSLDESKRK-----KLEEDINKLKTQLSFDLY--NKYKLERLFDKKKTVGK 1027

DB 814 EHQEDRKDKKKKGGHSSDKKEKYNKKEKTKESKSNILFDEEYIQLLELRTGECGFI 873

QY 1028 Y-KMOIKKLTLLKEQLESKLSL-----NNPKHVLQNSVFNNKKKEAE-TAETENTLEN 1080

DB 874 YLKSLSKELDVILNKSKDDALLNDAFNKINLAITSWNIFNEENKEGDNITVTENTATE 933

QY 1081 TKILLKHVYKLVKYNGESSPLKTLSESIQEDN-----YASLENFKVLSKLGK 1131

DB 934 GNI-----TIDENTEVMNEEYKIFSVEKIDMLAKEVGE 970

RESULT 34

ABG06301

ID ABG06301 standard; Protein; 2017 AA.

XX AC ABG06301;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #6292.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS70488.

XX PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity

XX PS Claim 20; SEQ ID No 36660; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity

CC . and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ

Query Match		5.1%	Score 425.5;	DB 22;	Length 2017;
Best Local Similarity		19.2%	Pred No. 2.2e-10;		
Matches 356;		Conservative 328;	Mismatches 678;	Indels 489;	Gaps 79;
QY	24	SYOELVKKLE-----ALDVAITGYSLSFKQKMLVINEGTSATVTTPSGSGS	72		
DB	163	SYRNLOQKEVEISHLKAQIALQDQ-----FLKLSAAQSPVSGAGVPATTASS--S	213		
QY	73	VASGGS-----GGSVASGGSVASGGSVASGGSVSGGSGNSRTPNSDSSD	119		
DB	214	FAYGISHPSAFHDDDFGDIITSSQOEINRLSNEVSRLESEVGHWRHIAQTSKAQGTDN	273		
QY	120	SDAKSYADLKHVRNYLLTIKELKY--POLFDLTNHLMTLCDNTHGFKY-----L	167		
DB	274	SDQSEICKLQ-----IINKLQNSQOEIDHQHMSVLQNAHQOKLFEISRHRREEL	326		
QY	168	IDGYEEINELLYKLNFYDLIRAKLNDVCANDYQIQIPFNKIRANELDVLKLVGYRKP	227		
DB	327	SDYEERTEEL-----ENLLQOGSGVETDLSKI-YEMQKTIQVLQIEK-----	369		
QY	228	LDNTKQNVGMEDYIKKNTKIENINELIESKTTIDKNKNAKEEKKILYQAOYLSI	287		
DB	370	-----VESTKKM-----EQLEDKIKDINK-KLSSAENDRDILRREQOLNV	409		
QY	288	YNKO-LREAHNL-----ISVLE-----KRDTLTKKNEI-----KELLDKINEI	325		
DB	410	EKRQIMEECMLKLECSKLOPSAVKQSDTTEKERIILAQSGSVEEVPRLQALSDAENEI	469		
QY	326	KNPPANGTNPNTL-----LDKNKKEEHEKEIKETIRFN-----IDS	367		
DB	470	MR---LSSNQDNLAEADNLKLMRIEVLKESLSLQKEELQMS-LLKLINYEVIKS	525		
QY	368	LFTDPLEYLLREKNKIDISAKVETKESTEPNEYPNGVTYPLSYDNDINNALNELSFG	427		
DB	526	TATRDISLDSLH-----DRLNLEAKE-----QELNQSISEKETLI	562		
QY	428	DLINPED-YTKPSKNIY-----TDNERKPFINEIKETIKTKKI---ESDKKSYE	475		
DB	563	AEIEELDRQOEATKHMILIKALQSKOONEGDSIISKLQDLNDEKRRKVHOLEDDRM---	619		
QY	476	DRSKSLNDITKEY-----EKLNEIY-----DSKFNNNIDLTNFEKMMGRKYSY	519		
DB	620	-----DITKELDVQFVLLIQSEVALNDLHLTKQKLEDKVENLVQDLN--KSOESNVS	670		
QY	520	KVEKL-----THINTF-----ASYENSKN-----LE	541		
DB	671	OKENLELKEHTRQNEEELSIRNEMQSLNODSNSEKFTLLKEREAEVRNLKQNLSELE	730		
QY	542	KLTKALAYME-DYSLRNIVVEKELKYKNLSKLTEN-----ELETIVENIKKDEE	590		
DB	731	OLNENLKKVAFDVKWN---EKLVLACEDVRHQLEECLAGNNQSLSEKNTIVETLRMEKG	787		
QY	591	QL-----FEKKTKDENKPDKEITILEYSDIVKVOQVKLL-----MNKIDELKKTQL--IL	638		
DB	788	EIEAELCWAKKRLAEANKYEKTEIELSARNALNTSALQLEHEHLIKLNQKDMIEAELK	847		
QY	639	KNVELKNIH-----VPSNYSQENKQBPYYILVILKKEIDKLKVPMPKVESLINEEKNK	693		
DB	848	KNIEQMDTDHKTEDVLSSSLEEQK-----LTQLINKKEIFEIKL-----KERSKK	894		
QY	694	TEGQSDN-SEPSTGEITGATTPGQAGSALGDSVQAQEQKQKQAPVPVPVPEAK	752		
DB	895	LQELDYSQALRKNELITRIETEEKDRSLGSKMEKNHLOEELERLREEQSRTPAVAD--	952		

QY	753	AQVETPPAPVNNKNTENYSKLDYLEKLYEFLMNTSYCHKYIL-----	793
DB	953	-----PKTLDVSVELASEVSQNLNTIKEHLEEEELKHQKILEDQNSQMQLLSIQEQKK	1006
QY	794	-----VSHSTMN-----EKILKQYKITKEESKLSLSCDPLDLL-----	826
DB	1007	EMDEFYHOEOMNATHQTOLFLEKDEEIKSLQKTIETQIKTQJHEERODIQTQNSDIFQETK	1066
QY	827	---FNIGNNIPVMYSMFDLSNLSOLFMEIYKEMVNCNLYKLDNDKDKIKNLEEAARKS	883
DB	1067	VQSLNIENG---SEKHDLSKAETERLVKGIKEREL-----EIKLNEKNISLT	1111
QY	884	TSVKTLSSSSMOPUSLTPQDKP-EVSA-----NDTSHSTNNLSKLFPENILSLGKKNKI	938
DB	1112	KQIDQLSKDEVGKLTQITQOKDLEIQALHARISSTSTQDV---VYLOQQLQAYAMEREK	1168
QY	939	YOELIGOKSSNFYEKILKDSDTFYNESFTNFVSKADDINSLNDESKRKKLEEDI---N	995
DB	1169	VFVNLNKTRENSHLK-----TEYHK-MMDIVAAKEAALIKLODENKKLSTRFESSGD	1221
QY	996	KLKKTQLSFDLYNKYKILKLERLFDKKKTV-----	1027
DB	1222	MFRETIONLSRIIEKDIIDALSOKCOTLLAVLQTSSTGNEAGVNSQFEELOERDK	1281
QY	1028	YKMOIKKLTLLKEQLESKLSNPNKHYLQNFVSFFNKKKKEAEIAETENTLTKILLKH	1087
DB	1282	LKQOVKMEEWKQOV---MTTVQNMQHESAQLOBELHQLAQVLVDSN---NSKLQV-D	1334
QY	1088	YGLVKYNGESSPLKTLSESIQIOTEDNYASLENFKVLSKLEGLKONLNKKKLSVLS	1147
DB	1335	YTLGIQSYEQNETKLKNFGQELAQVSHIGOLCNTKDL--LLGKL-DIISPOLSASLLT	1391
QY	1148	-SGLHLLIAELKEVIKNKNYTGNSPSPNNTDVNNALESYKKFPEGTDVAVTVWSESG---	1203
DB	1392	POSACLRAKSEVL-----SESSELQOELEELRKSLOEKDATIRTLQENHRL	1441
QY	1204	SDTL-EQSQPKKPASTHVGASNTITTSQNVDDV---DDVIIVIPGSESEDDDLGQV	1259
DB	1442	SDSTAATSELERKEHEQTDSEIKOLKEKQDVLOKLLKEKDLI-----KAKSD-----QL	1491
QY	1260	VTGEAVTPSVIDNLSKLENEYEVLYLKLPLAGVYVRSLLKQLENNVMTFNVVKDILNSRF	1319
DB	1492	LSSN-----ENFTNKV-NENELL-----QAVTNLKERILLILEMDIGKL-----	1529
QY	1320	NKRENFNV--LESDLIPYKDLTSSNVVVKDPYKFLNKEK-----RDKFLSSVNYIKD	1370
DB	1530	-KGENEKIVETRYCKETEYQALQETNM---KFSMWLEKEFECHSKMEKALAFEOELLE	1584
QY	1371	SIDTDFINFANDVLGYKILSEK---KSDLSISKYINDKOGENEKYLPLFNNIETLYKT	1427
DB	1585	KEQKKTGELNQLNAVKSMQEKTVVFQOERDOVMLALKOKOMENTALQNEVORLRDKEFR	1644
QY	1428	VNDKIDLVFIHLEAKVLNNTYKESNVEVKIKELNVLKTI-QDKL---ADFKNNNFVGIA	1483
DB	1645	SNOELERLNLHLESEDSYTRALAAEDREAKRKKTVTVEELKLVSSSNAMENASHQASV	1704
QY	1484	DLSTDYNNHLLTLKFLSTGMVFENLAKTLVSLNLLDGNLMLNISOHQVCKQCPQNSGC	1543
DB	1705	QVESLQQLNVVSKQORDETALQLSVQEOVK-----QYALSLANQMLVMEHQOEKA	1757
QY	1544	FRHLDERECKLLNYKQEGDKCVENP-----NPTCNENNGCDADAKTEE	1590
DB	1758	MYSAELEKQKQLIAEWKXKTKILEGKVISLQECLEDAANAALDSASRLTEQ	1808

RESULT 35
AAU00669
ID AAU00669 standard; Protein; 206 AA.
XX
AC AAU00669;
XX
DT 07-SEP-2001 (first entry)
XX

QY 1458 KELNYLKTQDKLADPKKNNNFGVGIADSLTDYNNHNNLLTKFLSTGKMFENLAKTVLSNLL 1517
DB 1936 -----NKNQIIG---YSIKYDKN-----VVSNSCSDVITSILK 1966
QY 1518 DGNLOGMLNISQHCVKKQPCNSGCFRHLDEREC-KCLLANYKQEGDKCVENP 1570
DB 1967 DKKIKR-----KKLQKRN---YENENIVCLDCLISYLKKMURIYGNP 2007

RESULT 38
AAB18195
ID AAB18195 standard; Protein; 1516 AA.
XX AAB18195;
AC AAB18195;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:52.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX Plasmodium falciparum.
OS Plasmodium falciparum.
XX WO200025728-A2.
PN 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
XX 05-NOV-1998; 98US-0107131.
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
DR
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 120-124; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are refined or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAB70078 to AAB70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 1516 AA;
SQ

Query Match 4.9%; Score 411.5; DB 21; Length 1516;
Best Local Similarity 21.3%; Pred. No. 6.5e-10;

Matches 363; Conservative 262; Mismatches 559; Indels 523; Gaps 92;

QY 141 ELKYP-QLFDTHNMLTCLDNIHGFYKL-IDGYEIN-----ELLYKLNFYDLRLAKLN 193
DB 37 ELTYANNVKDLNKSF---DNMSIFNDLWDFSENISSEIKTIDNIKKAHLYFFLR---- 89
QY 194 DVCANDYCOI-PF-----NLKTRANELDLVK---KLVFGY--RKPLD 229
DB 90 -ICKLLYNIPIFIFDGNPPPELKRRTIFQRIKKNRYEKEKFAEKLVVYQYORTL 148
QY 230 NIK-----DNVGMEDYIK-----KNKTIEN-----INELIEESK-----K 261
DB 149 SMKSNKNDNSNIEDTKNTNTQKNTQKSNTPNKINADISKSLIYDDIK 208
QY 262 TIDKNKATKEBEKKLYQAQYDLISYKQLEEAHNLISVLE----- 303
DB 209 EKDSLINSIVHGVNVPVSVKDVLTICNDLSKIKNKIFMITDFGVPVFLGQDGMGV 268
QY 304 ---KRIDLKKNENIKELLDKINEIKNPPPPANGTNPNTLLDKNK-IEHEKEKEIAK 359
DB 269 ENINKLDNRKNDENLSYINYNKVD---VNNNDDDR--DKKENINEVRDQKNYY 323
QY 360 TIKFNIDSLFDPLELYLREKN---KNIDISAKVETKESTEPNYPNGV--TV----- 409
DB 324 KNKENINNIYDDDEKEDIQNKNGVYNNDDIDEQIIRKKHMKRYYESIPKTFGFLC 383
QY 410 ---PLSYNDINNALNEL-----NSFGDLINPFDTYKPS-----KNI-YTDN 447
DB 384 MRPVDIIDSINYTEMLEISETLKVHENKFKOHLNVLDENNSTPVVNNMLLNINYNKN 443
QY 448 -----ERKKEFN-----ETKEKIEKKKIE-----SDKSY----- 474
DB 444 DDLIEGGEKKSPINLINVDSYSSNSRLNDENIETRGKINFMFITNDEKSIINNNNN 503
QY 475 -----EDRSKSLNDITKEYEKLNEIYDSKN-----NN 503
DB 504 NNNNNNNNDNDNDVIEHKNKNMNIYDNKYVCSSEKINDGINSKNNINIELPN 563
QY 504 IDLTNFEKMMGKRYSKYKVEKLTHTNTFASY--ENSKHNLEKLTALKVMEYSLRIVVE 561
DB 564 LDTSNI--FLECKDEYKYVYVKNKEIRIPLFKEINKEIFELKPLQYO---ILQDI--- 615
QY 562 KELYYKNLISKIEN-----EIEFLVENIKKDEQLFEK-KITKDEN--KPDEKI 608
DB 616 KEEWYTDNRKAIKSKDDMDVFSQVQLEYTVRMKTDFE--IEKLKMAENIQSVESGL 673
QY 609 LEVSDIVK---VQVQVLLMNMKIDELKKTQJLLKVELKHNHVPNSYK-----QENK 658
DB 674 LINKDLSKNTDNIINIKDYNVLOKKKKKKKKFLNDILNTYNTTESKYQDLYVKGESK 733
QY 659 QBPYYLIVLKKEDIKLVFMPKVESLINE--EKKNIKTEGQSDNSEPSTEGETQATTK 716
DB 734 ED-----IKNQIDFVTOECYRNNDIIRDTHDKSDFIKNIKIDNN-----K 773
QY 717 PQQAGSALGDSVQAQOEQAQPPVPVPEAKAQVTPPPAPVNNKT-----ENWSK 771
DB 774 KYEIVNLELEQEBEINEKKNYNNKNNND-----SNKTFELKIENEFK 813
QY 772 LDYLEKLEFLNTSVTICHKYILVSHSTWNEKTLKOYKITEEESKLSLSDPLDLFNION 831
DB 814 KDLL-----LDSQIFGDSLLAD-----IKENTYAD-----NLDN 844
QY 832 NIPVMSYMFDSLNNLSLSLFME-----IYKEMVCNLYKLKN----- 869
DB 845 -----NNENKSLIYEDGENFITRNEPITNEYEKNNIYISDEQYNEEDIIFK 892
QY 870 DKIKNLLBEAKKVSYVKTLSSSSQPLSLTPQDQPEV-SANDPTSHSTNLSNKLFPN 928
DB 893 DKIK---EKEKNNDTSSDDFENCYSVQEKIYVNEKEIEYNKNDKSSS---SSSILE- 945
QY 929 ILSLCKNNIYOELIGQKSENFEYKILKSDTFYNESTFNVKSKADDINLSNDESKRK 988
DB 946 -IKYKKEK--DELV-----SPNLC--VLLDEFEHNDLENNIYISSVDDMKTNVSKNNT 996

Pr falciparum and Candida albicans are used to detect infections of these
 PT pathogens in mammals -

PS Claim 10; Page 104-111; 138pp; English.

XX The invention relates to identification and use of telomerase
 CC reverse transcriptase (TERT) genes and proteins isolated from
 CC Plasmodium falciparum, Candida albicans and rice. TERT gene is
 CC also called as ever shorter telomere (EST) gene. TERT genes
 CC are used to detect P. falciparum and C. albicans infection in
 CC mammals. They are also used to diagnose the state of an infection
 CC in a patient and the relative amount of the pathogen in a cell,
 CC tissue, organ or organism. The potential binding partners and
 CC modulators of the activity of TERT genes and proteins are used
 CC to inhibit or promote growth of the pathogens and hence are used
 CC to treat e.g. stomach cancer, malaria and vaginal candidiasis in
 CC mammals.

CC The present sequence is Plasmodium falciparum TERT full length
 CC protein.

XX Sequence 2184 AA;

Query Match 4.8%; Score 402.5; DB 22; Length 2184;
 Best Local Similarity 20.4%; Pred. No. 2.5e-09;
 Matches 362; Conservative 268; Mismatches 585; Indels 557; Gaps 94;

Qy 114 SDNSDS--DAKSADLKHVRNYLLTIKELYPOLFDTNHLTLCDNIHG-FKYLIDG 170
 Db 148 NDNISDKCITTKNIPKLVHINKYKYLKK-KYHTMY--TNN-----DHSYGYLYLV-- 197
 Qy 171 YEEINELLYKUNFDDLRL-----AKLNDVCANDYCOQPPFNKIRANELDVLUKK 219
 Db 198 --QCSGRLLKNDFFKDMQIOERKKYTSNIAKINSEYTNNI--IINNNTNNNNNNNNNN 253
 Qy 220 LVFEGYKPLDNT-----KONVGMEDYIKKNKKT--IENINELIESKKTIDK----- 265
 Db 234 NVHGFGH--INNLFPSSNEPSSISCTWTETKNDKLTHTRETSLLITENSCKDKLLPEI 312
 Qy 266 -----NK-NATKEEKKKLYQAOYDLISYNKOLEE 294
 Db 313 DFSEDRKESVGYDKKXKXSNKIKRINKNTKEKKK-----WNKIIN 362
 Qy 295 AHNLLIS-----VLEKRI--DTLAKNENIKELDKINEIKRPPANGNPPTLLD 342
 Db 363 RNNILQHTNTKCKTFLFNKHIIFDKIENNIPLFIYDLLNYI-----FKSDQTYFY 414
 Qy 343 KNKITEEHEKEI-KEI-AKTKFNIDSFTDPLEEYV---LREKNKIDISAKVETKE 396
 Db 415 HNNFIDEYKQKICKOICKSTKNDLSHITSRKENHLPVOKLENNYKHPNINKOL---R 471
 Qy 397 STEPNEYPNGVTYPLSYNDINNALMELNSFGDLINPFDTYTRKPSNIYTDNERKKFNEI 456
 Db 472 KTKILKY---VYNYFKEFINNVIN--TKFGKIYRK-----FPRKHILNKI 512
 Qy 457 KEKIKIEKKKTESKKSYESDRSKSLNDITKEYEKLNLNIYDSKFNNDLNFNFKMMGKR 516
 Db 513 HKIFKIIRLQI---IKKYR-----IINRMNRKFIKQKYDT-FFKNYDFLSF----- 556
 Qy 517 YSYKVEKLTHTNTFASYE-----NSKHNEKLTALK-----YMEDYSLRNIVVEKE 563
 Db 557 -SFYTKIINFMVYITRKCIPIKLGSKHNFKIFLKNKYKFLFLPNYKESFSLNQVMKNIK 615
 Qy 564 LK-YKKNLISIKIENIEIETLVENIKKDEQLFEKTKTDENKPDKEKILVSDIVKVOVO-- 620
 Db 616 VKNIFQKKISKYINKNRLLKNIFDNN---YENKILHRNK--ELITWINNIKIYKNKN 670
 Qy 621 -----KVLNMNKI-----DELKKTQLIKKNVELKHNHVPNSYQENKQBPYLYIV 666
 Db 671 DNLNNSFKIKTTLNKLARKYFNKIKKINIAIQKRLANRL---IYFLFN---YFIMP 722
 Qy 667 LKKEIDKLVKVPKVESLINEEKNKIKTEGOSDNSEPSSTEGTEIGQATPKQGOAGSALE 726
 Db 723 LIRRF-----FFLTKSEQTLH---KTIFFDRKIWNH-----FTK-----IS 755

Qy 727 GDSVQAQAEQKQAPVPVPVPEAKAQVPTPPAVNNKNTENVSKLDYLELY-----E 780
 Db 756 NFCLYHQIFRNKKL-----KKRNEPKMDIVQNMFNVKKGE 791
 Qy 781 FLNTSYICHKYILV-----SHSTMNEKILKQYKITKEESKLSLSCDPLDLLFNI--QN 831
 Db 792 KIKT-----NKYIFIKMKKKSTNKICINNFKSKCIPKPKKKKN-----LYNITRHN 837
 Qy 832 NIPVMYSM-FDSLNSLSOLFMEIYEREMVNCNLVKLKDNDK-----IKNLEBAKVS 883
 Db 838 NIFIKKMEKSKTNNLIN-----KSIDNLKLEIKNKSVRPYIKKPYIYKIKKY 888
 Qy 884 TSVKTL-----SSSMQPLSLTPQDPEV-----SANDDTSHSTNLN 920
 Db 889 FALKKMYIHRMAKEEKSNIKLERAFKHFFIFAQEKELHILKYFSSHFFQNRKINYGKRFN 948
 Qy 921 NSLKLFENIL-----SLGKNKN-IYQELIGOKSSNFYEKILKDSOTFYNESFTNFVSK 974
 Db 949 KLIHRIKNIITIKQNSGIVKNKDKTFLHLTKNSKNNNK-KKNKNYNNNNNN--NNN 1005
 Qy 975 ADDINSLAND-----ESKRKKLEEDIN--KLAKTLQLSFDLYNKYKLERLFDKK 1023
 Db 1006 NNNNNNNNNNNCKLSNKRNYRNNNNNNKAKNENKNNIDSN-----LEK---KK 1057
 Qy 1024 TVGKYKMOIKKLTLLKEQLESKLSLNNPKHVLQNSVFFNNKKKEAEIAETENTLENTKI 1083
 Db 1058 KIYIVKI--KNIIEKRNFMKLNSINH-----FISKLRIN----- 1091
 Qy 1084 LLKHYKGLVYYNGESSPLKTLSESIOTEDNYASLENFK-----VLSKLEGLKD-- 1134
 Db 1092 WIPKKKGLRPLIN-----LSTLNVPFIVKQRIETELSKKSEFEYFHILNILEREKDK 1146
 Qy 1135 NLNLEKK--KLSYLSLSSLHLHILAEKLVKKNYTGNSPENNTDVNNALE--SYKKFLPE 1191
 Db 1147 NIKRRKKNKKNFNPVSLNNICNLSKLGCMRHHNNNSLFKMTLTKTGEIELEKLLKWL-- 1204
 Qy 1192 GTDVATVYSESGDTLEQSPKKPASTHVGAESNTITTSQNVDDDEVDDVIVPIPESEBE 1251
 Db 1205 -----HYLKNWFKKKRMKKYI--KNKLNKKKIYAVI-----CIGFESN 1242
 Qy 1252 DYDDLQGVVTEGAVTPSVIDNLSKIENEYEVLYLKLPLAGYVRSLLKQLENNVMT-FNVN 1310
 Db 1243 CYEHINHNLYLPK-ILKNFEDNI-----NNFEIYIL--FKRSFRLYNKLNLSFLSYYPVN 1294
 Qy 1311 VKD-----ILNSRFNKREN-----FKNVLESDLIPYKDLTSSNVVWDPYKF 1352
 Db 1295 VKSGLHVIRNLRELIIKSHLNDNHHLNOMFKTKSKDL-----YIFADSYKS 1344
 Qy 1353 LNKERKDKFLS-----SYNYI-----KDSIDTDINFANDVLGYIKILSEYK 1394
 Db 1345 LQVDRKIDFMTIITVIRYIYLYNIFYSIKFEKLNKRNIYFQFOENOMKGVY--LSVRDK 1402
 Qy 1395 SLDLSIKK-YIND-KQGENEKYLPPLNNIETLYKTVNDKIDLFTVHLAKVLNLYYEKSN 1452
 Db 1403 KRVENIKWYLNKMKKINHDEITLESKNSSI---NINKNFMICMTHQD-----TEERGN 1455
 Qy 1453 VEVKIKELNYLKTIO---DKLADFKKNNFVFG-IADLSTDYNNHNLTKFLSTGMVFEN 1507
 Db 1456 TONKEKHDIYIGPIYNNNSFDTSTTTTHSSNNYKGNINHVGDTYKNDGLLHGK----- 1506
 Qy 1508 LAKTVLSLLDGNLQGLMLNIHQHOCVKKQCPQNSGCFRHLDEREBCCKLLNAYKQBGDCV 1567
 Db 1507 -----NNSMNECYVKDI-----KCN 1521
 Qy 1568 ENPNTCTENNGCGDAD---AKCTEEDSGSN---GKKTCTCTKPSYPLPDGIFCSCSS 1620
 Db 1522 NNNNNNNNNNNNNNSYNKLCNVTNNSKNDIIVKHTIDTDSKNHTY--FKNKFL--- 1576
 Qy 1621 NFL-----GISFLLILMLIYSFI 1639
 Db 1577 NFLOKKIISNIYGLPQGFSLNLSLYAYIL 1608

RESULT 41

AAB18265
ID AAB18265 standard; Protein; 2013 AA.

XX AAB18265;

XX 07-NOV-2000 (first entry)

XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:122.

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

XX WO200025728-A2.

XX 11-MAY-2000.

XX 05-NOV-1999; 99WO-US26796.

XX 05-NOV-1998; 98US-0107131.

XX (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.

PA (GARD/) GARDNER M.

PA (VENT/) VENTER J C.

XX Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -

PS Disclosure; Page 285-291; 577pp; English.

CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.

XX Sequence 2013 AA;

Query Match 4.7%; Score 399; DB 21; Length 2013;
Best Local Similarity 19.5%; Pred. No. 3.2e-09;
Matches 366; Conservative 235; Mismatches 593; Indels 684; Gaps 85;

QY 157 LCDNTHGRKYLIDGYEENELL-----YKLNFFDLRLAK----- 191

DB 15 ISDNI--FCFLKDGVCIFCNLLNNEKYLITCSQDEGVAAQYVDFVVKCRVKEKDECN 72

QY 192 -----LNDVCAN-----DYCQIPFNKLIRNELDV 216

Db 73 KNMTINIMLLQNEKKIIKETCYIKNVVTNKIYHTFLVINKHYHNILCSLSFENNSEFI 132

QY 217 LKKLVGYRKPLDNIKDNVGMEDYIKKNKKTPIENINELIBESKKTIDKNKATKEEKK 276

Db 133 LN-----TNFVKTFGK-----IKSWACTNNNIFVLIKKKKLIINKN-----NQMS 176

QY 277 KLYOAYDLSYINKQ-----LEBAHNLISVL-----EKRIDTLKKE-----NIKEL 318

Db 177 KILNQNLVSKHTLDRSLLLMKGENDVNVICBSKKEKKKKKTKDNKNEKKKGHMEIKDV 236

QY 319 LDKINEIKPPANGTPTNLLDKNKKIEEHEKEIKETIKFNIDSLFTDPLELEY 378

Db 237 NEKINEKINE-----EKNEKINEEKNEKNEKIN-----EBSK 272

QY 379 LREKKNIDISAKVETK-----ESTEPNEYPNGVTPLSYNDINNALNELNSFGDLINPFD 434

Db 273 NEEKNEKINEEKNETNDKPYEEKENDNIPLDHSHVQYNIPTFSI--LNK----- 321

QY 435 YTKEPS-KNIYTDN-----ERKKFINEIKIEKKIESDKSYEDSKSLNDITKEYE 489

Db 322 --KEPDLKKIQFSNIILPIKKMIICPYDEKILII-----LLSHKSIIVIITNKNNDLKNMF 375

QY 490 KLLNEIYDSK-----FNNNIDLTNPEKMMGKRSYK-VEKLTHTHTF-- 530

Db 376 IIKELIFNSPIITTTWIDNYIFLIYFLANLFLSPAKPCRNLYFYKCNINNYSHITSFPY 435

QY 531 -----ASYENSKHNLE 541

Db 436 KSRNLYISFKTEIVCFKIRYVEIPLTVFKKIQTTEGVIDAKYLFERKPRVINTNHQS 495

QY 542 KL-----TKALKYMEDYSLRN-----IVVEKELK-----YKKNLSIKTIEIETLVENI 585

Db 496 NAKDDKGDNDVIREEDF--LRNNKNFSDVKRRKRNDKNYEIFNINILREIKTLENKI 554

QY 586 KDEQLF--EKKITDKENKPKDEKILEVSDIVKVOQVLLMKNIDEKTKTLLKNVEL 643

Db 555 SNNDYNIETDEGEINKDELKN-----RLSARSLSVYKNTFNLML-----L 594

QY 644 KHNHVPNSYKOEKQEPYIL-----IVLKKIEIDLKLVFMPKPVESLINEKNI----- 692

Db 595 GHNNKKKWIRODIRNMVHNKYNCVEEDVCINRYEKESIFV-----EYDNNNDNMLWS 650

QY 693 -----KTEGOSDNSEFSTEGETGQATTKPGCQAGSALGDSVQAQAOEQKQAPVP 745

Db 651 HLYFLKKKKKKFKEDPHYNDENVI-----KLLDFVSI----- 682

QY 746 VVPPEAKAQVPTPPAPVANKTENVSKLDYL-----EKLYEFLNLS--Y 786

Db 693 -----INLHKYILANNITSFYIWSKYLVFLDNLGLLYTKKNDGKIYDFLELSNFY 733

QY 787 ICH-----KYILVSHSTMN-----EKILQYKITKEE 814

Db 734 ICYKYNKINKVIDIKIINEHDIVYMDKKHLKKNLKNLYNLTINTKEKI-QSYNIF---- 788

QY 815 SKLSCDPLDL-----LFNIQNNIPWYMSFSLNLSLSQLPWEIYKEKEMVNCNLYKL 866

Db 789 SMLNCTCFIPLNDGSPFYFINITKHILLY---ENLQN-----FSNLGHNIQYCNFKKN 840

QY 867 KD-----NDKIKNLEEAQKYSTSVKTLSSSMOPLSLTPQDKPVSANDDTSHST 917

Db 841 KYIQSVFNKLNLEYIFNGFYVQOYIIFFFLIYSTSHKFFIYLVENITHIYLFKKIHQT 900

QY 918 NLNLSKLIFENILSLGKNKNIYQELIG---QKSSSENF-----YEKILKD-----SST 961

Db 901 NI-----LYKN---KEKNQONENIINMKROKESNYILYNYFYLYTKLNKDYVCLCSDK 952

QY 962 ----FY-----NESTNFV--KSKADDINSND-----ESKRKK 989

Db 953 SVSYFYMFDFLPREEEKMYISDKKKKINNSNDKNKYIYNRSGKNDNNYKENQKNE 1012

QY 990 LE-----EDLNK-----LKKT-----LQLSF-----DL 1007

Db 1013 VENHYDDDDDDNKSPYLYTRNIFFCSTKNTNIVYAKCIGNTMIIVADYLTNTFFYIKDN 1072

SQ • Sequence 2295 AA;		Query Match 4.6%; Score 388.5; DB 21; Length 2295;		Best Local Similarity 19.9%; Pred. No. 1.1e-08;		Matches 370; Conservative 282; Mismatches 591; Indels 615; Gaps 98;	
Qy	178	LYKLNFD--LLRAKLDVCAQYCIQIPFLKIRANELDVLKLVFGYRKPDLNIDNV	235				
Dy	5	VFLNFIFSFLLIUKDESVSNGRKEIYF-----DDEKLKLSFFDRS--TNINLDV	56				
Qy	236	G---KMEYI-----KKKKTIENINELIESKKTIDK--NKNATKEEKKLQAO	282				
Dy	57	GENDESSYVPREVDEKKKKKKIDS-----KENSNGNIYKNDTKNNE-----DYN	106				
Qy	283	YDLSTYNKQLEAHNLISVLEKRTDTLKKNEIKELIDKINEIKNPPANGNTPTLLD	342				
Dy	107	YNVVLKGRAREG---IITDEKRSSTKGRKNEONNNKMSDDVDHNNNMNDINFVE	163				
Qy	343	KNKIEEHEKEIKE--IAKTIKFNDISLTPDLELEYLYREK--NKNIDISAKVETKESTEP	400				
Dy	164	YKMTIDNYDKLDELILKLSINRNNYFN--MLDEYSLOTKLKEM-----	207				
Qy	401	NEYPNGVTYPLSYDINNALNELASFGDLINPFDYTKEPSKNI--YTDNERKKFI--NEIK	457				
Dy	208	-----YDSLNYLILMNN-----KNSRYFISFSNNEKKKIKNDMN	244				
Qy	458	EKIEK-----KKIES--DKKS---YEDRSKSLNDITKEYEKLILNEYDSKF	500				
Dy	245	ENIYRHFIVSLRWYNFKLIEFCFDKNFIYIDENK--IYSKYNYKMLNLFSSNF	303				
Qy	501	NNIDLTNFERK---MGKRYKYVEKLTH-----HNTFASYN---SK	537				
Dy	304	LYIYNSKFSLEIIDNYKYSFIINNIRKDPNNMYVCQSYDFIYSYFLSYNHHFDPK	363				
Qy	538	HNLEKTKALKYMEDYSLRNIWKE-----LYKYNLISKIENIETLVENIKKDEBOL	592				
Dy	364	H-----KYLINMDIWNNSIQTKQIGHKLYKLLKLNENLI--LYNYIKNDISEM	412				
Qy	593	FEKKTIDENKPEKILEVSDIVQVQVLLMKNKIDELKKTQILKNVELKHNHVPNS	652				
Dy	413	-----IPVTLERMIFSNF-----TNLLIDILNKLYNIDYQDN	446				
Qy	653	YKQEN-----KQPPYLIIVLKKEID--KLK--VFMP--KVESLINEEKNKIKTEGSD	699				
Dy	447	IKQENVNVNPQADAPQDVVHNKNDVSLKNVKEPKVEH--NKAMSNEYTERDGDMIYD	504				
Qy	700	--NSEPSTEGETI-----GOATTKPGQQAAGSALSGSVQAOAEQKQAO	741				
Dy	505	NTNKEFEKSEGTNNISGGEDTFKNISGGEDTFKNISEGDEYDGDG--EGDGDGDGEA	563				
Qy	742	PPVPVPPEAKAQVTPPPAPVNN-----KTENVSKLD--YLEKLYEFLNTSYICH	789				
Dy	564	DDSSVDTHNNKNDGKESDVNLLMDSYKKLANDENFKKYKYLKMLDFLNKSSSEKK	623				
Qy	790	KYILVSHSTMHEKILQYKTIKEESKSSCDPLDLAFNIQNNIPVMYMFDSLN--NSL	847				
Dy	624	EDI---NSYKNKYLKEGIIYNNKYSDY-----IPLIFNPTKDV-----FTSIQINIK	669				
Qy	848	SQI--PMEIYKEMVCLNYKLDKNDKIKNLEAKVSTSVKTLSSSSMOPLSLTPQDKPE	906				
Dy	670	SKINFFNIY--EYIITITKYKEN---KNFYODLLKCRREIFKDRHLLNENNI--MDKQE	722				
Qy	907	VSANDTSHSTNLNLSLKFENILSLGKNKI-----YQELIGQKSSSENFY--EKLTL	956				
Dy	723	-----ELKKNIRNLMRIHE--VSNEGNRRNTINRKYYKYTYDYDKMNELYYVEKNILN	774				
Qy	957	KSDTFYNESFTNFVSKADDINSLNDESKRKKLEEDINKLKTQLQSLFYLYKYLKL	1015				
Dy	775	VNDTNTF-----NFMNNEKDKNVF-----DIN---KTMRI--DYIYNNINLNI	813				
Qy	1016	-----ERLFPKKTIVGK---YKQIKLTLTLKEQLESKLSLNNPKHVQNF--SV	1061				
Dy	814	FTPAAIKMKDIYDQKILLRSNFVEKLKSNESICVLSFLYLIGINDDGKLFHPYGFPRNI	873				
Qy	1062	FFNKKAEAEIAETENTLTKILLKHVKLVKYNNGESSPLKT--LSEESIQTEDN----	1115				
Dy	874	DFSVKL---IREGDKGLCNFLSGVLYHINLPIFVNNSSISISTEMDDVLENDNSINSF	930				
Qy	1116	-----YASLENF-----KVLS-----	1126				
Dy	931	FYIYKNNENIRNHDLSDENRIIPRKEDNIKSIISYSLGSSKDDFFSKLAFTNNVIRL	990				
Qy	1127	KLECK-----LKDNLNLEKKLSY-----LSSGLHLHIAELKE-----	1159				
Dy	991	KYKNTNNTYLDKYDFDFTDKINKSVIKNNVSPFLTCDYLLSNILGAVVDSLRNSST	1050				
Qy	1160	-----VIKNKN-----VTGN-----	1169				
Dy	1051	LESVYEENINDKNNIIONTVQVQKNLFEYFVKLADNRNSYALAAELGEIYVLGNESTGI	1110				
Qy	1170	-----SPSENNTDVNNA-----LESYKKFLP-----EGDVATVVSBSG	1203				
Dy	1111	ERDEIKAFEFWKKAADQDGTTSALSTGYALDEYKFLKKEELVKNMDBREDILTMIHLEN	1170				
Qy	1204	SD-----TLQSQPKRPASTHVGAESNTIITSONVD--DEVDDVIIVIFGESEDIYDL	1256				
Dy	1171	STDKKNVTLFEMFOESSEKKKQKKKEKQDQDGTDRVDDKIVQVNVGVFOQSYGVN	1230				
Qy	1257	GO-----VVTGEAVTPS--VIDNLT-----SKIENEYEV---LYLKPLAGVYRSLLKOLEN	1302				
Dy	1231	DESMGRNGSIDCFMPPSGGLNYSVQNNANIONNANIONNANIONNANIONNANIONNA	1290				
Qy	1303	NVMT-----FNVNVKDILNSR--FNKRENFKNV--LESDLI-----PYKDLTSS	1342				
Dy	1291	NIQSNANIQSNANIQSNVNSHGNTNRQNNINNVFFENAVTQOTSYSGWANPSEDFVNN	1350				
Qy	1343	NYVVKDPKFKLKEKDKFLSSYNIKDSI--DTDINFAND-----	1381				
Dy	1351	SFSSVPSFSLFDIPEG---SEYEHTENILDEQNMFFTNKNKEQEQEGGNPNESNGMN	1407				
Qy	1382	-----VLGYKYILSEKYKSDLSIKYI-----NDKQGEN--EKY-----LPFLNNI	1421				
Dy	1408	DENDEMIKKYKMDLNDLNLKSLNAEYFHKAIARNDDSLLENILAKYNIHFKGLGTERNI	1467				
Qy	1422	ET--LY--KTVNDKIDLFVHLEAKVNLVITYEKSNEVVI-----KELNY-----	1462				
Dy	1468	ELAGIYLLKAADKGD-----NISQMLLGHYSGSDIGIKLNDYKDDDKIENLRKSYKYK	1523				
Qy	1463	-----LKTQDKLADFKNNFVGIADLSTDYNHN--NLTK	1497				
Dy	1524	SAQNGNIISLYNKSILILKGVNPKYKTFNKECKEKLKHFHFIGL-----FNERLYMLTK	1577				
Qy	1498	FLSTGMVPEN-----LAKTVLSNLIDGLQMLNISQHCYVKKQCPQ-----N	1540				
Dy	1578	LLRRNYQDKYTGSLLSLIMSEL--GDHAHNVNASMLWTLLKRTMQQFTEKYNIIVENLK	1635				
Qy	1541	SGCFRHLDEREECKLLNYKQGDKCVENPNTNENNGGDADAKCTEEDSGSGNGKK	1598				
Dy	1636	LSLIKELNKKEE-----KEKEKKNKNIHNVNNNNNNGYKKC--DKNCNDNVRK	1684				
RESULT 44							
ID	AAV31946	standard; Protein; 4134 AA.					
XX	AAV31946;						
XX	21-DEC-1999	(first entry)					
XX	Plasmodium falciparum	cytoadherence gene protein GLAG9 paralogue.					
XX	CLAG9; paralogue; cytoadherence linked asexual gene; CLAG;						
XX	erythrocyte; red blood cell; malaria; infection; therapy; vaccine.						
OS	Plasmodium falciparum.						

PN W09949048-A1.
XX 30-SEP-1999.
XX 25-MAR-1999; 99WO-AU00213.
XX 25-MAR-1999; 98AU-0002580.
XX (MENZ-) MENZIES SCHOOL HEALTH RES.
XX Kemp DJ, Trenholme KR, Gardiner DL, Holt DC, Cowman AF;
PI WPI; 1999-591099/50.
XX New proteins useful for treatment of Plasmodium infections in humans,
XX especially malaria
XX
XX Disclosure; Page 90-115; 150pp; English.
XX
XX This sequence represents the protein product of a clag9 paralogue
XX gene (see also AAZ20056) on chromosome 3 of Plasmodium falciparum.
XX It is a paralogue of the newly identified cytoadherence linked
XX asexual gene 9 (CLAG9) protein (see AAY31945) encoded by the clag9
XX gene on chromosome 9 of P. falciparum. CLAG9 facilitates
XX cytoadherence of erythrocytes parasitised with Plasmodium to other
XX cells. The identification of clag genetic species, and the
XX products encoded by them, enables a range of therapeutic agents to
XX be rationally designed and/or identified that are useful for the
XX prophylaxis and treatment of disease conditions caused or
XX exacerbated by infection with Plasmodium spp., e.g. malaria,
XX especially in humans.
XX
XX Sequence 4134 A4;
XX
XX Query Match 4.6%; Score 387.5; DB 20; Length 4134;
XX Best Local Similarity 21.1%; Pred. No. 2.4e-08;
XX Matches 357; Conservative 245; Mismatches 525; Indels 567; Gaps 93;
XX
XX 134 NYLLTITK-----ELKYPOLFDTNHLMTLCDNIHGKYLIDGIEINELLYKLNFY 184
XX 114 NYLNTVKVNSFSSWVELKEQQRERLFHYNMT-----KGVVSFIDSFEHKGPNTHICMV 168
XX 185 FLLLRALKNDVCAN--DYCOIPNL--KIRANEL-----DVLKLV 221
XX 169 FEFMGPNLLSLKHVDYKGIPLNVRKIAIHLVIGMOYLHDVCKIITHSDIKPENLVLSPL 228
XX 222 FGVRKPLDNKO-----NVCKMEDYIKKKKTI-----ENINELIESKKTIDKNKA 269
XX 229 TTIPKPRDYTKKLESNKSQVEKKENDQNDVKLLITTMNNINTNLSEKKVI-----ND 284
XX 270 TREERKKLYQAQYDLSYNNKQLEEAHNLISVLEKRIDTLKKNENIKELDKINEIKNPP 329
XX 285 TOKNDR-----NIEYDQKTSK-----ENIEDNVSFVNDPSDN 319
XX 330 PANSNG--TPNTLIDKNKKIEHE--KEIKETAKTIKFI-----DSLPT 370
XX 320 QKNLNNITDNNIIPSNVQIEKQSTLSKNKKNEKDSYININSLTNDQDLKREDIKEN 379
XX 371 DPLE--LEYLREKKNIDISAKVETKESTEPNGVYPLSYNDINNALNELNSFGD 428
XX 380 DXAEGITKYDMLNLIKNNISIKKINDCHS--PENKN-----KDNHQCED-----423
XX 429 LNPFDYTPKPSNIYDNERKFFNEIKEI-KTEKKIESDKKSYEDRSKSLNDITKE 487
XX 424 --NSINICNKNNTIOTNIND--NTVNEKINNTSKKDLMLANTQNNNDSEK--NDVVE 476
XX 488 YKLLNEIYDSFNNMIDITNEKMMGKRYSYKVEKLTHNT-----FASYENSKNNLEKL 543
XX 477 -QQLVNEIDLKKNKQ---TKKKKNINEP--PYVKHLRPSNSDPSLLTSYSIHAIQETL 531
XX 544 TRALKYMEDYSRLNIVVEKELAYKKNLISKIENIEITLVENIKKQEBQLEFKKITKDNK 603
XX 532 TRKPYHNTYFLNN--PEK-----YR-----DNKNMPLYLHRLPD-----CLKKIDQDDSD 575

QY 604 PDEKILEYSDIVKQVQVLLMKNKIDELKKTQILKLNVELKHNIHVPSNYSKQENKQEPY 663
DB 576 ETEEDDLSDVDQ-----NK-EQKNQL-----EVNLPNN-KYPNSNDVYK 614
QY 664 LIVLKEIDKLKVPMPKVESLINEE-----KNKITE-----GQSDNEP--- 703
DB 615 F--FEKDKINKFPIYCDMFNHLIHPALRHLHELYMKNKKNNIDSNMTNDLGNNSHKVY 672
QY 704 -STEGEITQATTKPGQQA-----GSALEGD-SVOAQAQSKQAQPPVPV 747
DB 673 INTE--DGEYCIPIYDFSVYHESKCYKIDGLSNLWIDESRYAEIQRQTRAEVIL- 728
QY 748 VPEAKAQVPTPPAPYNNKTENVSKLDYLEKLYEFLNTSYICHKYLIVSHSTM-NEKILKQ 806
DB 729 ----KSGF-----NETADI-----WSFACVMFELVTGDFLFPQ--KG 760
QY 807 YKITEESKLSKCDPLDLLFNQNNIPVMYSFSLNLSQLFWEIYEKEMVCNLYKL 866
DB 761 DRYDNEEH-----LSFIEVLGNIP--KHMDAGYNS-----HKYFNKNMYKL 802
QY 867 KDNDRKIK-----NLLEAKKVSTSVKTLSSSSMOPLSLTPQDKPEV----- 907
DB 803 KNIRIKKYGLYKILKYKNLPEK-----EISPLCSFLLPMLSDVPQTRPSAYTMLQHP 856
QY 908 -----SANDDTSHSTNLNLSKLFLNLSLGLKNNKIYOE----- 941
DB 857 WLMNYSLEEGDDMYINDESYISNDRNMKNNSNNFIYDGHNS-SKNKSNKKKKIDVN 915
QY 942 -LIGQKS-----SENFYEKILKSDIFYNESFTNFVKSADD-----INSLND----- 983
DB 916 YKIGNGNAYNDNYNKNYKNKN--NKNFNDVDVPEPSPQDMHANYNNDIVHAVLYEK 973
QY 984 -----ESKRKLEEDINKL-----KTKQLSFDLYNKYKLELFLDKKK 1023
DB 974 PYNSNNVISTYNNKGHKNFDINYLQHRNDNNSNQNISLTNTDFTNSDYLANMDH-- 1031
QY 1024 TVGKYKMQIK-----LTLKEQLESKL--NSLNKPKHVLO-----NFSVFNNKKKE- 1068
DB 1032 --DTYRKQIINKIPAHQISKLDGKNFAYNESIQVHDFQOYNEHDFEYKFNKRFHA 1089
QY 1069 -----AETAENTILE----- 1079
DB 1090 HHKEMKHNDYDEBEDEDEDDDEDYEDVDYDDDEYDEGEHDADEKNNNEKQ 1149
QY 1080 -----NTRKLLKHYKGLKYNGESSPLKTL--EESIQEDNYASLENFKVLSK----- 1127
DB 1150 EQOYGEKYNYEHYENMGY-----NKNIQOLSYTNNDNENFCETONIIYLQNKRDINF 1205
QY 1128 LEGKLKDLNLEKKKLSYLSLSSGL-----HHLIAE--LKEVIKKNKYTGNS--PSN 1174
DB 1206 KECTPRNNINKEIKSKDYQSSKVINQKONYWYKIKENTKLRHAKKQHYSNNNINKND 1265
QY 1175 NTDVNNALESYKFFPEGTDVATVYSESGSDTLEQSOPKKPASTHVGAESNTITTSQND 1234
DB 1266 NTNIMNQIDTQDQISKNLHDLST-----NNNDQK-----HGAJQKMHNEKTND 1311
QY 1235 DEVDVIVPIFGESEEDYDLGVVTGEA-----VTPSVI-----DNIL 1274
DB 1312 KPLNDEEL-IENRDQNVNKNCKVINKNSCAYTKWVSFFKTIPIIFFFLCLNEKVL 1370
QY 1275 SKIENEYEVLYLKLPLAGVYRSLLKQLENVMTFNVNVKIDILNSRFNKRBN--FKNVLES 1331
DB 1371 CSI-NENENLG-----ENKNENANVTNENLKNLNEYDNIEQLKSMIGN 1414
QY 1332 DLIPKDLTSSNVV-----KDPYKF-LNKEKDFE--LSSYN-----YIKDSI 1372
DB 1415 DEL-HKNLTILEKLLLESLEKDKLKYPLLKQGTQELIDISKFNKKNITDADSETIIPV 1473
QY 1373 DTDINFANDVLGYKILSEK-----YKSOL-DSIKKYINDKQENEKYLPFLNIETLYKT 1427
DB 1474 QSSF--HDIVYEHKLEQSTIEYNSDLSDIKKI-----FI--VRTL-KT 1515


```
Db 669 G E F F F D E V E K D E K M G E F F D E V E K D E K M G E N I F E I P K - - - - - K D D V E I E E T Y S E K 723
Qy 1015 L E R L F D K K T V G K Y K M Q I K K L T L L K E O L E S K L N S L N N K P H V L O N F S V F F N K K K E A I E T 1074
Db 724 M W E I D E K - - - - - I K D K D E P E K I H D E K K E V E F F L I A D K K E - - - - - 762
Qy 1075 E N T L E N T K I L K H Y G L V K Y Y N G E S - - - - - S P L K T L S E E S I Q T E D N Y A S L E N F K V L S K L E G 1130
Db 763 E N E D S N V E I - L N I D K N N F V F E N K E T F E I D E K V S K M N E E D F V Y E N N - - - - - E T F E - C E D I F L 816
Qy 1131 K L K D N L N L E K K L S Y L S G L H L I A E L K E V I K N K N Y T G N S P S E N T D V N N A L E S Y K - F P L 1189
Db 817 K R E D N D S E N E K - - - - - E I D E I G E V I N I G R Y H L N K N N S Y D D V H I L T H D F K N E L L 866
Qy 1190 P E G T D V A T V W S G S G T L E Q S O P K P A S T H V G A E S N T I T T S O N V - - - - - D D E V D D V I I V P I F G 1247
Db 867 I E K Y N V D N I C S D - - - - - D N I Y D G N I C D D N I Y D - - - - - G 896
Qy 1248 E S E E Y D D L - - - - - Q O V V T G E A V T P S V I D N I L S K I E N E Y E V L Y L K P L A G V Y R S L K K O L E N N V 1304
Db 897 D N I Y S G D N I Y G D N I Y S G - - - - - D N I Y S G - D N I Y S G D N I H S G D N I Y S G - - - - - D N I 941
Qy 1305 M T F E N V N V D I L N S R E N F K N V L E S D L I P Y K D L T S S N Y V V K D P Y K F L N K E K R D K F L S S 1364
Db 942 D D D N I Y D G N I N S - - - - - G O N V E N L K E H K I A V N E S E E T A Q D I K E Y - - - - - E K R D N E F T 991
Qy 1365 Y N Y I K D S I D T D I N F - - - - - A N D V L G Y Y K I L S E K Y K S D - - - - - L D S I R K Y I N D 1406
Db 992 - D Y V E E - N S D I R F Y D K G K G E M V N E L I G E Y - - - - - S E K Y M D N N I E D N E L Y I W S A S V K - - - - - N D 1042
Qy 1407 K O G E N E K Y L P F L N N T I T Y - K T V N D K I - - - - - D L F V I H L E A K V - - - - - L N Y T Y E K S - - - - - N 1452
Db 1043 K E R L A D D N I D L N N N I S N D Y I K N N E D I K N V H D S F S I S K S E L H D I N G L E K S I S S N D I K S 1102
Qy 1453 V E V K I K E L N Y L - - - - - K T I O K L A D E F K K N N F - - - - - V G T A D L S T D Y N H N N L T K F L S T G M V E N 1507
Db 1103 I E V C V K E N E I H H K M M K K E L N D N N L N D E M Y M C D I S N D I F K N N E Y T K H V D D V Y T F D - 1161
Qy 1508 L A K T V L S N L D G - - - - - N L O G M L N I S O H Q C Y K Q C P Q N S G C F R H L D E R E C K C L I N 1558
Db 1162 - - - - - E N N S N L I I G E D E H C V S S M N F E Y P F N I S - - - - - K M T E S N N I L Y E Q N D K - - - - - 1205
Qy 1559 Y K Q E G D K C V E N P N P C N E N N G C D A - - - - - D A K C T E - - - - - 1589
Db 1206 - K K T I N S V K H P M Y I K G F A S D S I N E L K A L K G L P P L P F L K C K D M K P M R L F N I V L K V I 1264
Qy 1590 E D S G S G K K I T C E C T K P D S Y P L F D G I F C S S N F L G I S F L L I L M L I L Y S F I 1639
Db 1265 E S N D Y N G R R - K I K V T K M - - - - - F I C L K U K F D M I Y V F I I Y I L Y I F L 1305
```

RESULT 47

AAB18234

ID AAB18234 standard; Protein: 2573 AA.

XX

AC

AAB18234;

XX

07-NOV-2000 (first entry)

XX

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:91.

XX

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

XX

KW antimalarial; malaria; protozoacide; infection; insecticide.

XX

OS Plasmodium falciparum.

XX

WO2000025728-A2.

XX

11-MAY-2000.

XX

05-NOV-1999; 99WO-US26796.

XX

05-NOV-1998; 98US-0107131.

PR

XX

(HOFF/) HOFFMAN S.

PA

(CARU/) CARUCCI D.

PA

(GARD/) GARDNER M.

PA

(VENT/) VENTER J C.

XX

Hoffman S, Carucci D, Gardner M, Venter JC;

PI

WPI; 2000-365347/31.

XX

Proteins encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum, useful as antimalarial vaccines and in the diagnosis of P.falciparum infection -

PT

Disclosure: Page 213-220; 577pp; English.

PS

The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, Plasmodium falciparum. Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against P. falciparum infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against P. falciparum infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with P. falciparum. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent P. falciparum infection, or they can be used to identify drug resistance in P. falciparum. Sequencing of the plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic life cycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAB70078 to AAB70287 and AAB18144 to AAB18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

XX

SQ Sequence 2573 AA;

Query Match 4.6%; Score 383.5; DB 21; Length 2573;
Best Local Similarity 18.7%; Pred. No. 2.1e-08;

Matches 359; Conservative 288; Mismatches 625; Indels 643; Gaps 86;

Qy

119 DSDAKSYADLKURVRNLLTIKELKYPQLFDLTNHLMTLCDNIHGFYKYLIDGYEINELL 178
Db 189 ENKKDEKEKKNIHNN-----DENNNMI-YKKNIDKTHYILD--NNVVHL 232

Db

179 YKLNPFYF---DLLRAKLNDVCANDYQIPFNKIRANELDVLKLVGYRKLPLDNKDN 234
Db 233 NDINTYLRERDYMNRKFGTYIDSTY-KNPMYVTVLYIFNNDILKDIL---QVIDIIRD 288

Db

235 ----VGKMEDYIKKKTKTIENINEL-IEESKKTID--KNKNATKEEK-----KLYQAO 282
Db 289 FDHAIYKDIDENOLIKNLIILNHLTTRPSKEWFYWKRMPTFNDKSEYNYKYLOQ 348

Qy

283 -----YDL-----SIYNKOLEEAHNLISVLEKRIIDTLLKKNENIKELLKINEIKNPPAN 332
Db 349 KSDRILYDTLKNDIYIKELQRSIDLOYQKGLQSLKCLLANKNFLTMLNFY-----403

Db

333 SGNTPTLL-----DNKKIEE---HEKEIKEIAKTIFKNIDSLTFDPLELEYLR 380
Db 404 -----NTQLFDADYREITEENEKVMQORRENELLEEKRLK-----QELESYHD 448

Qy

381 EKNKNIDISAKVETKESTEPNEYPNGVYPLSYND-INNALNELNSFGDLINFPDYTKEP 439
Db 449 DSSTDDSSADQDQDERE-----VLTHNDPINKKDDPIN-----490

Qy

440 SKNIYTDNERKKFINEIKIEKKIESDKSYEDRSKSLNDITKEYEKLLNEIYDSK 499
Db 491 -KNDDPINKNDNDINKNDNI-----NKDNDICNSDHTNSNDHTNSNDHTNSNDHT 545

Qy

500 FNNIDLTNFEK---MMGKRSYKYEKLTTH-----HNTFASYENSKH-----NLEKLT 545

(1) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. AAY70078 to AAA70287 and AAB18144 to AAB19352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

Sequence 2206 AA:

Query Match 4.5%; Score 379.5; DB 21; Length 2206;
Best Local Similarity 19.9%; Pred. No. 2.6e-08;
Matches 294; Conservative 214; Mismatches 451; Indels 515; Gaps 72;

QY 240 DYIKNNKKT IENINEL-----I ESKKTI DNKNKNATKEEKKKLYQAQYDLSIYNKOLEEA 295

Db 2 EYLESEKSSDDRRREVNFFENDYSKDSHSHNSDLDVDRKKHSDNVYESEODGKOTEG 61

QY 296 HNLISVLEKRIDTLKKNENIKELLDKINEIKNPPPPANSNGTPTNTLLDKNKKIEEHEKEIK 355

Db
62 RKKI---KGFFKLLKKGDS-----|::|:|:|:|
EDENKEKETKDHRLE

QY 356 EIAKTIKENIDSLETFDPLEYYLREKKNKNIDISAKVETKESTEPNEYPNGVTYPLSYND 415

[illegible]

QY 416 IN----NALNELNSFGDLINPFEDYTKEPSKNIYTDNERKKKFINEIKEKIKIEKKKIESDK 471

```

Db      125 KNKHNGSSSNEHSSYSD-----ENFFAAKRNKILNE-----EYKKNKDNE 167

```

QY 472 KSYEDRSKSLNDITKEYEKLLEIYDSKFENNIDLTNFE-----KMMGKRY---SYKV 521

```

Db      168 MC--DMSIFNDD-----NNMDDSLFNKKNEDNNRYDEEEMKKYKRRKGKRYSSDSYKD 216

```

QY 522 EKLTHHTFASYENSKHNLEKLTALKYMEDYSLRNIVVEKELKYKKNLSKIENEIETL 581

```

Db      217 DSPQYMS--ERYSSSEKYSSEKYSSESSN-----RQSTNLLNNIKNFCNTY 268

```

QY 582 VENIKD-----EEQLFEKKITKD-----ENKPDEKILEV--SDI-VKVO 618

```

Db 269 I I K K K D R S R D T Y E D E S R E G A Y G E N T E D L N E D T Q E G H K N K K K E I L M N I L Y N D I N I K K N 328

```

QY 619 VQVLLMN-----KIDELKKTQLILKN-VELKHNIHV-----PNSYKQENKQEPYYL 664

```

Db      329  DEKDFEMDRNFKGKKKEIDIKKNQQVMKNMLNIKNNENIDVYNDKDNFINIDDKCPSGYF 388

```

QY 665 IVLKEIDKLKVFMPKVESLINEEKNI-KTEGQSDNSEPSTEGETTGQATTKPGQAGS 723

Db 389 -----DKDKIKEY-----DYLDNEKQKNVVKMIHPKDGNNNNNNNILLSQ----- 427

QY 724 ALEGDSVQAQAEQKQAPPPVPVPEAKAQVTPPAPVNNKT--ENVSKLDYLEKLYEF 781

Db 428 -----NSSTILSHVVQEDYADGIKFF 448

QY 782 LNTSYICHKYILVSHSTMN-EKILKQYKITKEESKLSSCDPLDLLFNQNNIPVMYSMF 840

Db 449 NKNSFYNN--LENRKLININNIYDKYKI-----ILSEIKSG-----482

Qy 841 DSLNLSQLFMEIYEKEMVCNLYKLKDNDKIKNLEEAKKVSTSVKTLSSSSMQPLSLT 900

Db 483 DHLNNVEKKL--NVIENSLLC-IFDLQDGDNNNNINDD----- 517

Qy 901 PQDKPEVSANDDTSHSTNLNNSLKLFFENILSLGKNKNIYQELIGQKSSSENFYEKILKDSD 960

```
Db 518 --DDNDDGGHDDGGHDDHVGNNMK-----GDK-----LDIK 546
```

QY 961 TFYNESFTNFVKSADINSL--NDESKRRKKLEEDINKLKKTQLSFDLYNKYKLLER 1017

21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.
09-JUL-2000; 2000US-0598042.
19-JUL-2000; 2000US-0620312.
03-AUG-2000; 2000US-0653450.
14-SEP-2000; 2000US-0662191.
19-OCT-2000; 2000US-0693036.
29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AA157837.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX Example 3; SEQ ID NO 1826; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 1780 AA;
XX
XX Query Match 4.5%; Score 375; DB 22; Length 1780;
XX Best Local Similarity 18.9%; Pred. NO. 3.le-08;
XX Matches 291; Conservative 267; Mismatches 563; Indels 422; Gaps 60;
XX
XX 130 HVRNYL-----LTIKELKYPOLFDTNH-----MLTLCDNHGFXYLIDG 170
XX 196 HRSREYLRSSQKEETASKSALLRQIKVEYTHNDSDDTLGYSUTNSLITSEFEESIKD 255
XX
XX 171 YEEIN-ELLYKLN-----YFOLLRAKLVNDVANDYQIPFNKIRANELDVLKLVFY 224
XX 256 YEQANLNMANSTKFSVWVSFFIYNEYIDLF-----VPVSSKQKRMRLSODVKGY 309
XX
XX 225 R--KPLDNIKDNVGRMEDY-----IKKNKKTITENIELIEESKKTIDKNKNATKEEEKK 276
XX 310 SFIKDLQWTOVSQSK-EAYRLKLGKIKHQSVAFTKLNNASSRSHSFTFKLIQIEDSEMS 368
XX
XX 277 KLYQAQYDLSTYN-----KOLEAHNL-----ISVLEKRIDLTKNE----- 313
XX 369 RVIRVS-ELSLCDLAGSERTMTKQNEGERLRETGNINTSLTLGRCINVLKNSKSKFQQ 427
XX
XX 314 -----NIKELL-----DKINEKNPPPPANSNGTNTLLD 342
XX 428 HVPFRESKTHYFQSFNGKGIKIMVINSQCYLAYDETNLVNLKFSATAQKVCVPDTLNS 487
XX 343 KKKKIEEHEKEIKETIKFN--IDSFTPTPLEYILREKNKNIDISAKVETKESTEP 400
XX 488 SOEKLFGPVKSSQVSLDSNSKILNVKRAATISWNSLDEDLMEDEDLVEELENAEETQ- 546
XX 401 NEYPNGVTPLYSNDINNALNELNSFGDLNPFDTKPSKNI--YTDNERKKFFINEIKE 458
XX 547 -----NVETKLLDELDKTLLENKAF-----ISHEEKRKLLDLIEDLKKKLINEKKE 593
XX

QY 459 KIKIEKKIESDKKSYEDRSKSLNDITKEY-----EKLLEIYDSKFNNDIL 506
DB 594 KLTLEFKIRE-----EVTQFTQYWAQREADFKETLLQEREILEENAERRL 639
QY 507 TNFKMMGKRYSKYKVEKLTHNTFASYENSKHNLEKALKYMEDYSLRNIVVEK-ELK 565
DB 640 AIFKDLVGKCD-----TREEAAKDICATKVETBEATACTACLELKFQKAEIAKTGELI 692
QY 566 YYKNLISKIENIEITLVENIKKDEQLEFKITTKDENKPDKEKILEVSDIVKQVKVLLM 625
DB 693 KTKELKKRENESSLIQLETSNKKI-----ITQNG-----RIKELINIIDQREDTI 740
QY 626 NKIDELKKTQTLILKNVELKHNHVPNSYKQENKQBPYLIIVLKKKEIDKLKVPMPK-VESL 684
DB 741 NEFQNLKS-----HMENTFKNDKADTSSLIINNKLICNETVEVPKDSKSK 786
QY 685 INEEKKNTKTEGQSDNSPSTEGEITQATTKPGQOAGSALLEGSDVQAQOQKQAOPPV 744
DB 787 ICSEKRVNENELQODEPPAKKGI-----HVSSAI-----TEDOKKSEYR 828
QY 745 PVPPEAKAQVPTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHYILVSHSTWNEKITL 804
DB 829 P-----NIAEIEDIRVLOE-----NNEGL 847
QY 805 KOYKITKEEESKSSCDPLD---LFNIQNNIPVMYSMPFDSLNLSLSOLFMEIYEKEMVC 861
DB 848 RAFLTLTIENELKNEKEKAELNKQIVHFQOEL---SLSEKKNLTLSKEVQOIQSNYDIA 903
QY 862 ----NLYKLKONDKTKNLEEAKKYSTSVKTLSSSMQPLSLTPDQKPEVSANDOTSHST 917
DB 904 TAEHVOKSKNOEQEIKMKLSNEIETATRSITNNVSO-IKLMHTYKIDELRTLDSVSQIS 962
QY 918 NL-----NNSKLFIENILSLGKNKIYQELIGOKKSENFY-----952
DB 963 NIDLLNRLDLSGSEBDNLPTQDILLGNDYLVSKQVEYR--IQENRENSFHSSIEAI 1020
QY 953 ----EKILKSDSTFYN-----ESFTNFVKRADDINSLSNDSKRRKLEEDINKLK 999
DB 1021 WEECKEIVKASSKSHQIEEQQIEKLAQAEVKGKVDENNRLKE--KEHKNODDLKKE 1078
QY 1000 TL--QLSFDLYNK---YKLKLERLFDKKKTVCK---YKMOIKKL-TLLKEOLESKL 1046
DB 1079 TLIIQTLKEELOEKKNVTLQVQHVVVEGRALSELTOGVTYKAKIKETILUETOKVERS 1138
QY 1047 NSLNNPKHVLQNFVSFFN-----KKRAEIAETENTLE 1079
DB 1139 HSAKLEQDILEKESILKLERNLKEFQEHLODSVKNTRDNLVKELKLKE-EITQLTNNLQ 1197
QY 1080 NTKILLKHYKGLVKYNGESSPLK-TLSEESIQT-----BDNYASL-----E 1120
DB 1198 DMKHLQLKEEBEETNRQETKLEELSASSARTQNLADLQKBEDYADLKEKLTDAKK 1257
QY 1121 NPKVLSKLEGLKKNLNLKELKLSYLSGLHLHIAELKEVIKNKNTGNSPENNTDNN 1180
DB 1258 QIKQVQKESVNRDEDKLLRIKINELEKKKNQCSOELD--MKORTIQQLKEQLNNQKVEE 1315
QY 1181 ALESYKKFLPEGT-----DVATVVSSEGSTLQSQPKKPASTHV-----GAE 1223
DB 1316 ALQOYERACKDLNVKEKIIEDMRMTLEEQEQTVQEQDQVLEAKLEVERLATELEKWK 1375
QY 1224 SNTITTSQNVDDEVDVIVPIFGESEEDYDDLQGVVGTGEAVTPSYVIDNLSKIENEYEV 1283
DB 1376 CNDLETKNRNSN-----KEHENNTDVLGKLTN-----LQDELOESEQKYNA 1417
QY 1284 LYLKPLAGVYRSLLKOLENNVMTF-----NVNVDIL-----NSRFKNRKNFKNVLES 1331
DB 1418 -----DRKKLEKMMMLITQAKEAENIRNKKMYAEDRERFPQONEMELITLA 1466
QY 1332 DLIPYKDLTSSNYVVKDPYKFLNKEKROKDFLSYN-YIKDSITDIDINFANDVLGYKILS 1390
DB 1467 QUTE-KD-----SDLOKW--REERDOLVAALQLKALISSNVQKDNIEOLKRIIS 1515


```
Qy* 433 #DYTKPSKNTYTDERKKFNEIKIEKIKIEKIESDKSYEDRSKSLNDITKEYEKL 492
Db 163 FNLEK-LKKNLGSGNSNENILN-----DSQKIENDQN-TNLSKEKS-----ENIL 207
Qy 493 NEIYDSKFNNDIDLTNEK-----MMGK---RYSYKVEK-----LTH 527
Db 208 KTPDSKYSNNNTTSLKKISSNSOKSELSPPSQTIGKIYRPSYLKELYEILDDI 267
Qy 528 NTFASYENSKHNLEKTKALKYMEDYSLRNIVVEKELYKN-LISKIENIEITLVENIK 586
Db 268 NT-GRVTLGKNRLKELKKGSLNFKQVKNELIENSKKEASNLLTLTKKDIEPNLINIP 326
Qy 587 KD--EEOLEFKKITYDENKPEKILEVSDIVKVOQVLLMNKIDELKKTOLILKNVE-- 642
Db 327 KDPYKKEIFQ--LQEDKPK-----QVLEDLKSQVHSIKPID-LENTKSROQAIDKLN 378
Qy 643 LKHNIHPNSYKQENKQOEPYLVILVKEIDKLKVPMPKVESLINEEKKNIKTEGQS--DN 700
Db 379 LKNN---PND-AQASK-----TLAQANKIOHLEDLKSQVHSIKPIDLENTKSROQAIDK 429
Qy 701 SEPTGEITQOATTKPCQQAQSALEGDSVQAQEQQAQPPVPVPEAKAQVPTPPA 760
Db 430 NEFLKNPNDAQA-SKTLAQANKIOHLEDLKSQVHSIK-----PIDLENTK----- 474
Qy 761 PVNKNTEVSKLDYLEKLYEPLNTSYCHKYILVSHSTMNEKILKQYKTKKEE--SKLS 818
Db 475 -----SRQAIDKIDNEFLKNN-----PNDQAASKTLAQANKIOHLEDLKSQVH 517
Qy 819 SCDPLDLLFNQNNIPVMSFDSLNLSOLFMEIYEKEMVCNLYKLKND-KIKNLE 877
Db 518 SIKPIDL--EN-----TKSROQAIDKIDNE-----FXKNPNDAQASKTLA 555
Qy 878 EAKVSTSVKTLSS--SMOPLSLTPQDKPEVSANDDTSHSTNLNLSLKFENILSLGKN 935
Db 556 QANKIQ-HLEDLKSQVHSIKPIDLENTKSROQAIDKIDNEFXKNPNDAQASKTLAQANKI 614
Qy 936 KNITQELIGQKSSNFYEKILKSDSYNESFTNFVSKKADDIN--SUNDESKKKLEED 993
Db 615 QH-----LED-----LKSQVHSIKPIDLENTKSROQAIDK 644
Qy 994 INKLKKT----LQSLFDLYNKYKLEKRLFPDKKTKVCKYKMQIKKLTLLKQLESKLSNL 1049
Db 645 LNEFLKNPNDAQASKTLAQAN--KIOHLEDLKSQVHSIKPIDLENTKSROQAIDKIDNEF 702
Qy 1050 --NNPKHVQLNFSVFFNKKAEIAETENTLEN---TKILLKHYK-GLVKYNGESSPL 1102
Db 703 XKNPNDAQASKTLAQAYENNGDLLKAENAYEKIILKLTQEDHYKLGIRF-----KL 756
Qy 1103 KTLSESIQTDNVALENFVKLSKLEGLKDNLNLEKKLSYLSGSLHHIAELKEVIK 1162
Db 757 KKYEH-----SIESFDQTLKDPKHK-----KALHNKGIALMMLNK 792
Qy 1163 MKNYTGNSPSENNTDVNNALESYKFFLPGCTDVATVVSSEGSDFLEOSOPKPPASTHVA 1222
Db 793 NK-----KAIESFEKAIOIDKNYGTAYQKG-----IAE 821
Qy 1223 ESN-----TITTSQNVDDVDVVIPIFGESEDDYDLGQWVTGEAVTPSVIDNIL 1274
Db 822 EKNGDMQQAFAFNKAYNLD-----KNPNY-----ALKAGIVSNL 857
Qy 1275 SKIENEYEVLYKPLAGYRSLKQLENVNMTFNNVKD-----ILNSRFNKRNFKNVL 1329
Db 858 GNF-----KOSEYLNFNANAKKPNIEAIYNLSIAKFEN--NKL 895
Qy 1330 ESDLIPYKOLTSNNVYVVDKPKFLNKEKRDKFLSSYNIKSDIDTDFNFANDVLGYKIL 1389
Db 896 E-----ESLETINKAID-----LNPEK-----SEYLYLKASINLKNENYQNAISLSLV 939
Qy 1390 SEKYSKLDLSIKKYN-----DKQGE-----NEKLYPLFANNIETLY---KTVN 1429
Db 940 IEK---NPENTSAVINLAKAYEKGSKNSQAISTLEKILNKNKLANLNLGILLYKKEKNYQ 996
Qy 1430 DKIDLFVHLEAKVLYNTYE--KSNVEVKIKELNLYKLTQDKLADF---KKNN-----NFVG 1481
```

```
Db 997 KAIEIF---EKALINSIEAKYNLATTLIEINDNTRAKDLLREYTKLPNNPEALHALG 1052
Qy 1482 IADLSTDYNNHN-----LLTKF 1498
Db 1053 I-----IETNENNDOTLRELKIF 1072

RESULT 54
AAY19934
ID AAY19934 standard; Protein; 1119 AA.
XX AAY19934;
XX 19-JUL-1999 (first entry)
XX B. burgdorferi antigenic protein, f742.aa.
XX Antigenic protein; vaccine; Lyme disease; infection; detection.
XX Borrelia burgdorferi.
XX WO9859071-A1.
XX 30-DEC-1998.
XX 18-JUN-1998; 98WO-US127118.
XX 03-SEP-1997; 97US-0057483.
XX 20-JUN-1997; 97US-0050359.
XX 22-JUL-1997; 97US-0053344.
XX 22-JUL-1997; 97US-0053377.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMMUNE INC.
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI; 1999-189980/16.
XX N-PSDB; AAX61631.
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the diagnosis, prevention and treatment of diseases
XX caused by Borrelia, particularly Lyme disease
XX Claim 12; Page 124-125; 275pp; English.
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX can be used in vaccines for eliciting protective antibodies to members of
XX the Borrelia genus, particularly for the use against Lyme disease in
XX humans and animals. They can be used for preventing or attenuating an
XX infection caused by a member of the Borrelia genus. The products can also
XX be used for detection of members of the Borrelia genus.
XX SQ Sequence 1119 AA;

Query Match 4.4%; Score 369; DB 20; Length 1119;
Best Local Similarity 21.3%; Pred. No. 3.3e-08;
Matches 286; Conservative 208; Mismatches 472; Indels 378; Gaps 68;

Qy 261 KTDKNKNATKEEKKLYQAQYDLSIYNKOLEEAHNLSVLEKRIDTLKKNENIKELLD 320
Db 33 KLNDKNEIMLNEVKN-----SVIDRNYKRAYSVAKLLQ---DKYPONEDIAMLTN 80
Qy 321 KINIKNPPPPANS-----GNTPTLLDNKNKTEEHEKEIKE---TAKTIKFNIDSFTDPL 373
Db 81 TLAIANSPPESKDLQDSANQILDKIGODNTKTNVNFDAFNRYIKDSTITE-- 138
Qy 374 ELEYLREKKNKIDISAKVETKSTENY-PNGVTYPLSYNDINNALNELNSFGDLNP 432
Db 139 --NYSRNDVDVGIEDIESEFKKIKPEIKPN--TNPKEEDQIIOQSPNPKSLVNDQKNL 194
```



```

Qy 1484 DLSTDYNNHLLTFLSTGMVFENLAKTVLSNLLDGNLQGLMNI SQ-----HQC VKKQC 1537
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2247 KTKQRNNKEKLKRSISLPL---NLKRTVVKII---NLKKNINLKNIIIDAINNDILKGT 2300
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 1538 P-----QNSGCPRHLDERECKC-----LLNYKQEGDKVCVEN 1569
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2301 PYEHIYHSNFWIYSSDESDFYNCSDYYLNNVDHGSKEFDN 2343
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 57
ABB63519
ID ABB63519 standard; Protein; 2346 AA.
XX
AC ABB63519;
XX
DT DT
XX
DE DE
XX
KW Drosophila melanogaster polypeptide SEQ ID NO 17349.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI: 2001-656860/75.
XX
DR N-PSDB; ABL07622.
XX

New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -

Disclosure; SEQ ID NO 17349; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2346 AA;

Query Match 4.3%; Score 363; DB 22; Length 2346;
Best Local Similarity 20.0%; Pred. No. 1.5e-07;
Matches 278; Conservative 240; Mismatches 525; Indels 344; Gaps 57;

Qy 171 YEEINELLYKUNFYFDLLRAKNDVCANDYCOIPFNLRANELDVLKLVFGYRKPLDN 230
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 395 YAKSSELEMRNCEIQKLQKLSIAIESAP-----ILEKQNSDYOK-MKE 442
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 231 IKDVGKMEYIKNKKTIENINELIESKTTIDKNKNATKEEKKKLYQAYDLSIY-- 288
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 443 TNSELREHDELLQNKLCLE---RELERALSTLHNQN-----ENKKLKQTHTDLSROVC 494

```


Db 1479 RYKDFYIGLSGGGTESAAELEKVRSELEEVNQLRALDKDEHEKITCEDEVKK---R 1534
QY 1220 VGEASTITTSNVDDVDVIPIFGESDYDDLGQVTVGEAVTPSVIDNILSKIKEN 1279
Db 1535 TPEPDTSTAIRQYKAKDKLVLDLTAVT-----DLVNOETTFAGTKSSYDETARLEK 1589
QY 1280 EYEVLYLKPLAGYRSLKKQL--ENNVMPTFNVN-VKDIILNSFRNFKNFVLESDLIPY 1336
Db 1590 E-----LQENIAANKDINQRLTRENESLHMRINQLTRQLGSOOSTKPSSTSSVAERGNISE 1644
QY 1337 KDLTSSN 1343
Db 1645 SSPRTAN 1651
RESULT 58
AAB18205
ID AAB18205 standard; Protein; 1712 AA.
XX AAB18205;
XX 07-NOV-2000 (first entry)
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:62.
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX Plasmodium falciparum.
XX WO200025728-A2.
PD 11-MAY-2000.
XX 05-NOV-1999; 99WO-US26796.
PF 05-NOV-1998; 98US-0107131.
PR (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX Disclosure; Page 141-146; 577pp; English.
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA/0078 to AAA/0287 and AAB18144 to AAB18352 represent nucleotide

CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 1712 AA;
Query Match 4.3%; Score 362.5; DB 21; Length 1712;
Best Local Similarity 20.0%; Pred. No. 1.le-07;
Matches 317; Conservative 262; Mismatches 574; Indels 429; Gaps .81;
QY 204 PF---NLKTRANELOVLKLVFGYGRKPLDNIDKNVKGKMDYIKKKNKKTINELIEESK 260
Db 20 PEYERNLNDKGHLKELKRL---ERVDEPLYNNDVKI-----PNKKEIYNNIKSNNIQ 71
QY 261 KTDKNKNAKKEE-----KKLYQAOYDLSIYNKQLEEAHNLISVLE--- 303
Db 72 VRVQNNEEKKKEBANYTCVNKKVYVTLKNKVHVNYK---VNNSNINKI-KIVPIIKCSN 127
QY 304 -----KRDTLAKN-ENIKELDKINEIKNPPANGNTPNTLLDKNKKIEBEHEKEIKEI 357
Db 128 YKIKNNPISHLKSNYENKFKLSNFSNIKN-GCSHKDNVINETMDQHKSEQLNNDNIKKL 186
QY 358 -----AKT-IFKNIDSLSLFTDPLEYLYLREKNKNIDISAKVETKESTEPNEY 403
Db 187 LDYDIFREDTITKTNTNISYNNKMSFKDNEENINYM--DNNNIKSSSSYCSYSKNINO- 243
QY 404 PNGVTPPLSYNDINNA-----LNELNS-----FGDLINPFDTYTKPSKNIYTDN 447
Db 244 -----NNVHTLKTTEFLNEKNSHTQEQSLPLDGLQNNHNSATKPHNNIYDNN 293
QY 448 ERKKFINEKEK- IKIEKK--KIESDKKSYEDRSKSLNDITKEYEKLKLNIEYDSKFN--- 501
Db 294 --NSLVNYKSKDGIDLHNKMMKIETDKNGIITLEKKKHD-----EKYYNINIEPLNLDNS 346
QY 502 NNIDLTNFEKMMGKRYSYK--VEKLTHHTTFASYENSKHNLEKLTAKALYMEDYSLRNIV 559
Db 347 NNVVITTCNKESYRNSTSDMINKI-----FEKMMNEKKNILK----MKNFMDVTKKKT 397
QY 560 VEKELKYKNLISKIENETETLVENIKK-----DEEQLEFKKI-----TKDENKPDCK 607
Db 398 MAKE-----KILNSNSTI--NMKKVSYFNSKDEDLFNEKENSYKYGVRKNOEDIN 446
QY 608 ILEVSDIVKQVQVLLMKNKIDELAKTQL-----ILKNVELKHNIHVNSYKQENK 658
Db 447 VIK-----NNMKRNINIDNNDNIINIKNDSYKNIHI--NNKKKRD 487
QY 659 QEPYVL---IVLKEIDKLK--FMPKVESLINEEKNKIKTEGQSDNSPSETEIGQA 713
Db 488 DPPFNNSAGLLDFDLCKRKKVLEIKNVQS--SKKKKILLTWHNHSNQNCHSSDNQNC 545
QY 714 TTKPQQQAGSA-----LEGDSYQAOAQEQKQAPPPVPVPPEAKAQVPTPPAP 761
Db 546 HSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDNQCHSSDN 593
QY 762 VNNKNTYNSKLDYLEKLYEFLNTSYICHKY ILVSHSTWNEKTLKQYKIT--KEESKSLSS 819
Db 594 MNKNNSNKKK-----NKRTEKSKKIKNNNNNNNNNNNN 625
QY 820 CDPL-DLLFNIOQ---NIPVMYSMFDLSNLSLSQLEMEIYEKEMVCNLYKLKDNDIKN 874
Db 626 CDMGDQRIINNMDKQNVNQNQNGFNNNKNNNDLLNVYISPNNIN-HSLSSCEKKN 684
QY 875 LLEEAQKYSTSVKTLSSSSMOPSLTPODKPEVSANDDTSHSTNL--NNSKLFLNLSIG 933
Db 685 --KEDKNKNDN-KFLNSSSKMKI-----PEISTNSNEKIVNSDNEMLYHNLTVLN 734
QY 934 KKNKIYQELIGOKSS-----ENFYKILKDSOTFFNESTNFKVSKADDINSLNDES 985
Db 735 VKE---QGGVTBESSCIKRTYFVDQFY-----DSYNNRNEKITD-DNMQVEDIYVVKENI 785
QY 986 KRKLEEDINKLTKTLOLSFD-----LYNKYKLEKLERLF----- 1019
Db 786 KR-TLKGOGHDDVKTNMLSEDSNYSAGLWNGNFINISNNENCLNSYDISCDEKYPNEEE 844

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 07:18:27 ; Search time 67 Seconds
(without alignments)
5040.466 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKIIFFLCFLFFIINTQCV.....SNFLIGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

SPTREMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mmc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5283.5	62.7	1704	5	Q9T2T4
2	5162.5	61.3	1694	5	Q9NHX1
3	5162.5	61.3	1694	5	Q9T2T5
4	5107.5	60.6	1720	5	Q25922
5	3361	39.9	656	5	Q8T6A9
6	3093.5	36.7	1751	5	Q26194
7	3003.5	35.7	1087	5	Q25961
8	2951	35.0	1726	5	Q02569
9	2801	33.3	539	5	Q9TYG1
10	2798	33.2	539	5	Q25972
11	2796	33.2	539	5	Q25966
12	2795	33.2	539	5	Q25973
13	2794	33.2	539	5	Q25976
14	2791	33.1	539	5	Q25981
15	2789	33.1	539	5	Q25984
16	2788	33.1	539	5	Q25971

Q25645	plasmodium	1787	5	Q25645
Q9U6D4	plasmodium	1791	5	Q9U6D4
Q25685	plasmodium	1785	5	Q25685
Q25668	plasmodium	1766	5	Q25668
Q25924	plasmodium	651	5	Q25924
Q25923	plasmodium	652	5	Q25923
Q03999	plasmodium	400	5	Q03999
Q9BMG8	plasmodium	376	5	Q9BMG8
Q9nat3	plasmodium	360	5	Q9NAT3
Q25975	plasmodium	569	5	Q25975
Q25969	plasmodium	569	5	Q25969
Q25977	plasmodium	569	5	Q25977
Q25974	plasmodium	569	5	Q25974
Q25979	plasmodium	569	5	Q25979
Q25983	plasmodium	569	5	Q25983
Q25967	plasmodium	569	5	Q25967
Q25968	plasmodium	570	5	Q25968
Q25982	plasmodium	569	5	Q25982
Q25970	plasmodium	569	5	Q25970
Q9TYG2	plasmodium	570	5	Q9TYG2
Q25978	plasmodium	569	5	Q25978
Q25980	plasmodium	569	5	Q25980
Q25865	plasmodium	336	5	Q25865
Q9T2U5	plasmodium	356	5	Q9T2U5
Q9T2U2	plasmodium	350	5	Q9T2U2
Q9T2U7	plasmodium	357	5	Q9T2U7
Q9T2U8	plasmodium	357	5	Q9T2U8
Q9T2V2	plasmodium	360	5	Q9T2V2
Q9T2U9	plasmodium	363	5	Q9T2U9
Q9T2U1	plasmodium	344	5	Q9T2U1
Q9T2V3	plasmodium	362	5	Q9T2V3
Q9T2V1	plasmodium	365	5	Q9T2V1
Q9T2U4	plasmodium	351	5	Q9T2U4
Q9T2U3	plasmodium	351	5	Q9T2U3
Q9T2V0	plasmodium	360	5	Q9T2V0
Q96448	plasmodium	353	5	Q96448
Q9T2V6	plasmodium	347	5	Q9T2V6
Q9T2V5	plasmodium	347	5	Q9T2V5
Q9T2U6	plasmodium	352	5	Q9T2U6
Q9T2V4	plasmodium	347	5	Q9T2V4
Q9T2T8	plasmodium	344	5	Q9T2T8
Q9T2T9	plasmodium	344	5	Q9T2T9
Q25760	plasmodium	363	5	Q25760
Q9T2W5	plasmodium	369	5	Q9T2W5
Q9T2W1	plasmodium	372	5	Q9T2W1
Q9T2V9	plasmodium	402	5	Q9T2V9
Q9T2W3	plasmodium	378	5	Q9T2W3
Q9T2W7	plasmodium	357	5	Q9T2W7
Q9T2U0	plasmodium	345	5	Q9T2U0
Q9T2T7	plasmodium	345	5	Q9T2T7
Q9T2W8	plasmodium	354	5	Q9T2W8
Q9T2W4	plasmodium	348	5	Q9T2W4
Q9T2W6	plasmodium	342	5	Q9T2W6
Q9T2W2	plasmodium	355	5	Q9T2W2
Q9T2V8	plasmodium	367	5	Q9T2V8
Q9T2V7	plasmodium	367	5	Q9T2V7
Q9T2W0	plasmodium	343	5	Q9T2W0
Q25724	plasmodium	373	5	Q25724
Q25725	plasmodium	372	5	Q25725
Q25721	plasmodium	373	5	Q25721
Q43997	plasmodium	372	5	Q43997
Q25726	plasmodium	372	5	Q25726
Q9NCM9	plasmodium	614	5	Q9NCM9
Q25722	plasmodium	373	5	Q25722
Q25723	plasmodium	373	5	Q25723
Q9NCM1	plasmodium	595	5	Q9NCM1
Q9NCM4	plasmodium	588	5	Q9NCM4
Q9NCM2	plasmodium	588	5	Q9NCM2
Q9NCN2	plasmodium	599	5	Q9NCN2
Q43995	plasmodium	373	5	Q43995
Q25719	plasmodium	372	5	Q25719
Q25720	plasmodium	372	5	Q25720
Q25727	plasmodium	373	5	Q25727

90	1173	13.9	372	5	Q25717	Q25717 plasmodium
91	1171	13.9	219	5	Q9UB87	Q9UB87 plasmodium
92	1170.5	13.9	373	5	Q43996	Q43996 plasmodium
93	1170.5	13.9	599	5	Q9NCM5	Q9NCM5 plasmodium
94	1168.5	13.9	373	5	Q25728	Q25728 plasmodium
95	1168	13.9	372	5	Q25718	Q25718 plasmodium
96	1165	13.8	239	5	Q9NGK6	Q9NGK6 plasmodium
97	1164.5	13.8	597	5	Q9NCM3	Q9NCM3 plasmodium
98	1164.5	13.8	597	5	Q9NCM6	Q9NCM6 plasmodium
99	1160.5	13.8	597	5	Q9NCM0	Q9NCM0 plasmodium
100	1159.5	13.8	595	5	Q9NCM0	Q9NCM0 plasmodium
101	1158.5	13.8	597	5	Q9NC657	Q9NC657 plasmodium
102	1157.5	13.7	597	5	Q9NCN1	Q9NCN1 plasmodium
103	1150.5	13.7	599	5	Q9NCM7	Q9NCM7 plasmodium
104	1147.5	13.6	613	5	Q9NCN0	Q9NCN0 plasmodium
105	1147	13.6	616	5	Q9NCM8	Q9NCM8 plasmodium
106	1105	13.1	213	5	Q9TY76	Q9TY76 plasmodium
107	1105	13.1	213	5	Q9TY75	Q9TY75 plasmodium
108	1105	13.1	213	5	Q9TY73	Q9TY73 plasmodium
109	1105	13.1	213	5	Q9TY72	Q9TY72 plasmodium
110	1105	13.1	213	5	Q9TY70	Q9TY70 plasmodium
111	1099	13.0	213	5	Q25928	Q25928 plasmodium
112	1099	13.0	213	5	Q25939	Q25939 plasmodium
113	1034	12.3	211	5	Q25707	Q25707 plasmodium
114	1021.5	12.1	636	5	Q26182	Q26182 plasmodium
115	992.5	11.8	214	5	Q25927	Q25927 plasmodium
116	992.5	11.8	214	5	Q25930	Q25930 plasmodium
117	977.5	11.6	214	5	Q25942	Q25942 plasmodium
118	973.5	11.6	214	5	Q25936	Q25936 plasmodium
119	968.5	11.5	214	5	Q25943	Q25943 plasmodium
120	954.5	11.3	214	5	Q25926	Q25926 plasmodium
121	954.5	11.3	214	5	Q25931	Q25931 plasmodium
122	954.5	11.3	214	5	Q25934	Q25934 plasmodium
123	954.5	11.3	214	5	Q25937	Q25937 plasmodium
124	954.5	11.3	214	5	Q25938	Q25938 plasmodium
125	939.5	11.2	214	5	Q25929	Q25929 plasmodium
126	939.5	11.2	214	5	Q25932	Q25932 plasmodium
127	939.5	11.2	214	5	Q25933	Q25933 plasmodium
128	909	10.8	215	5	Q25935	Q25935 plasmodium
129	909	10.8	215	5	Q25940	Q25940 plasmodium
130	827.5	9.8	218	5	Q9TVG8	Q9TVG8 plasmodium
131	689.5	8.2	379	5	Q25659	Q25659 plasmodium
132	686.5	8.1	144	5	Q25992	Q25992 plasmodium
133	659.5	7.8	138	5	Q25960	Q25960 plasmodium
134	656	7.8	121	5	Q9UAI7	Q9UAI7 plasmodium
135	656	7.8	121	5	Q9TYE4	Q9TYE4 plasmodium
136	655	7.8	121	5	Q9UAI6	Q9UAI6 plasmodium
137	649	7.7	119	5	Q9TYE7	Q9TYE7 plasmodium
138	645	7.7	121	5	Q9UAI8	Q9UAI8 plasmodium
139	642	7.6	327	5	Q26043	Q26043 plasmodium
140	637	7.6	121	5	Q9TYE6	Q9TYE6 plasmodium
141	637	7.6	312	5	Q00796	Q00796 plasmodium
142	634.5	7.5	315	5	Q00794	Q00794 plasmodium
143	634.5	7.5	315	5	Q00798	Q00798 plasmodium
144	634.5	7.5	315	5	Q00803	Q00803 plasmodium
145	634	7.5	116	5	Q9TYE3	Q9TYE3 plasmodium
146	633.5	7.5	315	5	Q00799	Q00799 plasmodium
147	631	7.5	312	5	Q00797	Q00797 plasmodium
148	631	7.5	328	5	Q9TYF2	Q9TYF2 plasmodium
149	630	7.5	328	5	Q9TYF0	Q9TYF0 plasmodium
150	628.5	7.5	315	5	Q00802	Q00802 plasmodium
151	628	7.5	338	5	Q9TYE9	Q9TYE9 plasmodium
152	628	7.5	338	5	Q9TYF1	Q9TYF1 plasmodium
153	628	7.5	338	5	Q9TYF3	Q9TYF3 plasmodium
154	626.5	7.4	339	5	Q00792	Q00792 plasmodium
155	625	7.4	304	5	Q00801	Q00801 plasmodium
156	625	7.4	338	5	Q9TYF4	Q9TYF4 plasmodium
157	624	7.4	338	5	Q00791	Q00791 plasmodium
158	624	7.4	338	5	Q00793	Q00793 plasmodium
159	623	7.4	338	5	Q9TYE8	Q9TYE8 plasmodium
160	619.5	7.4	328	5	Q9GSQ9	Q9GSQ9 plasmodium
161	617.5	7.3	339	5	Q00800	Q00800 plasmodium
162	616.5	7.3	333	5	Q00795	Q00795 plasmodium
163	610.5	7.2	2771	5	Q26216	Q26216 plasmodium
164	606.5	7.2	1939	5	Q25662	Q25662 plasmodium
165	568.5	6.7	2269	5	Q26223	Q26223 plasmodium
166	559.5	6.6	1979	5	Q96133	Q96133 plasmodium
167	558.5	6.6	2747	5	Q9BJX9	Q9BJX9 plasmodium
168	558	6.6	2867	5	Q9N2M3	Q9N2M3 plasmodium
169	556	6.6	3130	5	Q9BK46	Q9BK46 plasmodium
170	554	6.6	3254	5	Q9BK45	Q9BK45 plasmodium
171	553	6.6	96	5	Q9TYE5	Q9TYE5 plasmodium
172	553	6.6	2752	5	Q9BJY0	Q9BJY0 plasmodium
173	539	6.4	105	5	Q9TY71	Q9TY71 plasmodium
174	528	6.3	95	5	Q964N2	Q964N2 plasmodium
175	525	6.2	95	5	Q964N3	Q964N3 plasmodium
176	524	6.2	105	5	Q25941	Q25941 plasmodium
177	521.5	6.2	3394	5	Q77384	Q77384 plasmodium
178	518.5	6.2	308	5	Q02560	Q02560 plasmodium
179	517.5	6.1	308	5	Q02619	Q02619 plasmodium
180	516.5	6.1	308	5	Q02620	Q02620 plasmodium

ALIGNMENTS

RESULT 1

Q9TZT4

ID

Q9TZT4

PRELIMINARY;

PRT;

1704 AA.

AC

Q9TZT4;

DT

01-MAY-2000 (TREMBlrel. 13, Created)

DT

01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT

01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE

Merozoite surface protein 1.

GN

MSPI.

OS

Plasmodium falciparum.

OC

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX

NCBI_TaxID=5833;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=HN2;

RA

Jiang G., Liu R., Daubenberger C.A., Pluschke G.;

RT

"Sequence analysis of the MSP 1 gene of Plasmodium falciparum from Hainan, China."

RL

Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).

DR

EMBL; AF062349; AAC72885.1; -

DR

InterPro; IPR000561; EGF-like.

DR

Pfam; PF00008; EGF; 1.

SQ

SEQUENCE 1704 AA; 193763 MW; 385526D0DA56FDID CRC64;

Query Match

62.7%;

Score 5283.5;

DB 5;

Length 1704;

Best Local Similarity 62.1%;

Pred. No. 2.4e-157;

Matches 1075;

Conservative 219;

Mismatches 317;

Indels 121;

Gaps 26;

QY

1

MKTIIFLCFLFFIINTQCVTHESYQELVKKLEDAVLITGYSLFQKMKVINEGTS

60

Db

1

MKTIIFLCFLFFIINTQCVTHESYQELVKKLEDAVLITGYSLFQKMKVINEGTS

60

QY

61

AVTSTPGSK---GSVASGGSGSVAGSGSVAGSGSVAGSGSGSRRTNPSDNS

117

Db

61

AVTSTPGSKGSGSVAGSGSGSVAGSGSVAGSGVA---SVASGGSGSRRTNPSDNS

117

QY

118

SDSDAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHLMTLCDNIHGFYKYLIDGYEEL

177

Db

118

SDSDAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHLMTLCDNIHGFYKYLIDGYEEL

177

QY

178

LYKLNFFDRLAKLNDVCANDYCIQIPFNLKIRANELDVLKLVFGYRKPDLNKNV

237

Db

178

LYKLNFFDRLAKLNDVCANDYCIQIPFNLKIRANELDVLKLVFGYRKPDLNKNV

237

QY

238

MEDYIKNNKTTININELIEESKKTIDKNNATKEEKKLYOAYDLSYNNKLEAHN

297

Db

238

MEDYIKNNKTTININELIEESKKTIDKNNATKEEKKLYOAYDLSYNNKLEAHN

297

QY

298

LTSVLEKRDITLKKNNIKELDKINEKNPPANSNGTNTLLDNKKITEEHEKEI

357

Db

298

LTSVLEKRDITLKKNNIKELDKINEKNPPANSNGTNTLLDNKKITEEHEKEI

357

Db 298 LISVLEKRIDTLKKNENIKLLDKINEIKNPPANGSNTPTNLLDKNNKIBEHKEIKEI 357
Qy 358 AKTKFNIDSLFTDPLEYLYLREKNKNDISAKVE--TKESTEN-EYPNGVYPLSYN 414
Db 358 AKTKFNIDSLFTDPLEYLYLREKNKNDVTPKSDPTKSVQPKVPYNGVYPLPLT 417
Qy 415 DINNAL---NELNGFGLDINFDYTKPSKNIYTDN-ERKKFNEIKKIKIEKKI--- 467
Db 418 DIHNSLAANDKNSYGLMNP-DTKEKINEKIITDNKERKIFINNIKKQIDLEEKNIHT 476
Qy 468 -ESDKSKYEDRSKSLNDITKEYEKLLNFIYDSKNNNIDLTFEKMGKRSYKVEKLTH 526
Db 477 KEONKKLLDEYKS---KKDYEELEKFEYEMKNNNDKVDVDFISARYTYNWKORY 532
Qy 527 HNTFASYSNKHNEKLTALKYKALYEDYSLRNIVVEKEKLYKYNLISKITENIEIETLVENIK 586
Db 533 NKKFSSNNNSYVQKALKALYEDYSLRGISEKDPNHYITLTKGLEADIKKITEIK 592
Qy 587 KDEOLFPEKK---ITKDNKDPDEKILEVSDIVKQVQVQLVLMNKIDELKTKTOLILKNVEL 643
Db 593 SSENKILEKFKGLTHSANAS---LEVSIVKLVQVQVLLIKKIEDLTKLTELFLKNAQL 648
Qy 644 KHNTHPNSYKOEKQEPYIIVLVLKKEIDKLKVPMPKVESLINEEKNKIKTEGOSDNSEP 703
Db 649 KDSHTVPNIYKPNKEPEYIIVLVLKKEIDKLKKEFTPKVKMDLKKKEQAVLSS----- 699
Qy 704 STEGEITQATTKPGQOAGSALGDSVQAQOEKQAO-----PPVPVPVPEAKAQVPT 757
Db 700 ITQPLVAASETDEGGHSTHLSQSGETEVEETEETETVGHYTTVTITLPTQ---PS 756
Qy 758 PPAPV-----NKKTENVSCKDYLEKLYEFLNTSYICHKYILVSHSTMEKILKQ 806
Db 757 PPKEVKVVENSTEHSKNSDNOALTKVYLKLDLFTKSYCHKYILVSNSSMDQKLEEV 816
Qy 807 YKITEESKLSQDPLDLNIONNIPVMSYDMSDLSNLSOLFMEIYKEMVCNLYKL 866
Db 817 YNLTPKEENELKSCDPLDLNIONNIPAMYSLYDSMMNDLQHLFFELYQEMYYLHL 876
Qy 867 KDNOKIKNLEAKV-----STSVKTLSSS-----MQPLSLTPTQDKPVS 908
Db 877 KEENHIKLLKEEQKQITCSTSSPGNTVTNTAQSAATHSNQONASNTQNGVAVS 936
Qy 909 AND---DFTSH---TNLNSLKLFENTLSLKGNNKIYQEL-IGKSSSENPEYKILKSD 960
Db 937 SGPAVVEESHDPVLVLSISNDLKGIVSLNGLNKTVPNPLTISTTEMEKTYENILKND 996
Qy 961 TFYNESFTNFKSRADDINLNDSESKRKLLEEDINKLAKTLQLSFDLYNKYKLERLFD 1020
Db 997 TYFNDDIKQFYKSKSVITGLT-ETQKNALNDEIKKLKDTLQLSFDLYNKYKLERLFD 1055
Qy 1021 KKKTVGKMKQIKKLTLLKEQESKLSNPNKPVLFQNFVFFNKKKEAEIAETENTLEN 1080
Db 1056 KKKELGQDKQIKKLTLLKEQESKLSNPNKPVLFQNFVFFNKKKEAEIAETENTLEN 1115
Qy 1081 TKILLKHVGLVYNGYNGESSPLKTLSEBSIQTEDNYASLENPKVLKLEGLKDNLNLEK 1140
Db 1116 TKILLKHVGLVYNGYNGESSPLKTLSEBSIQTEDNYANLEKFRVLKIDGKLDNLDHLGK 1175
Qy 1141 KKLVSLSGLHLHLTAELKEVTKNNYTGNSPSENNTDNNNALESYKFKPLPGTQVATVVS 1200
Db 1176 KKLVSLSGLHLHLTELKEVTKNNYTGNSPSENNTDNNNALESYKFKPLPGTQVATVVS 1234
Qy 1201 E-----SGSDTLEQSQPKKPASTHYGAESNTITTSQNVDDVDVITVPIF 1246
Db 1235 PPQDPTVPSLVSVRVSGSGGSKTEQIPTSGSLTELEQQVQVQSONYDEEDSLVVLPIF 1294
Qy 1247 GESEEDYDLQGVVTGEAVTPSVIDNLSKIENIEVEVLYLPLAGVYRSKLKQLENVMT 1306
Db 1295 GESEDNDEYDLQGVVTGEAVT-NDNLSGFENEYDVLYLPLAGVYRSKLKQLEKNIT 1353
Qy 1307 FNVNVDILNRSFNKRNKFNKVLSDILPYKDLTSSNVVVKDPYKFLNKRKDFLSSYN 1366
Db 1354 FNLNLDILNLSRLKRRKVFLDVLESDLMQFHHISSNEYIIEEDSKFLNSEQNTLLASYK 1413

Qy 1367 YIKSIDTDFINFANDVLGYKILSEKYSKSDLSIKKIY-----NDK 1407
Db 1414 YIKSEVNDIFAEGISYIEKVLAKYKDDLESIKKVIKEKEFPSSPTTPPSPAKTDE 1473
Qy 1408 OGENEKYLPPFNNIETLYKTVNDKIDLEVIHLEAKVLANYTEKSNVEKIKELNLYKTIQ 1467
Db 1474 QKSEKFLPFTNITETLYNVLNVNIDDLINLKAINDCNVEKDEAHVKITKLSOLKAIT 1533
Qy 1468 DKLADFKNNNFVGADLSTDYNNHNLTKFLSTGCMVFENLAKTVLSNLGDLNLOGMLNI 1527
Db 1534 DKIDLFRNTDFEATKRLINDDTTKDMIGLKLSTGLV-ONFPNTIISKLEIGFODMLNI 1592
Qy 1528 SQHCVKKQCPQNSGCFRHLDERECKCLLANKQSDKCVENPNPTCNENNGGCDADAKC 1587
Db 1593 SQHCVKKQCPQNSGCFRHLDERECKCLLANKQSDKCVENPNPTCNENNGGCDADAKC 1652
Qy 1588 TEEDSGSNGKIKTCBCTKPDSPYPLDGIPTCSSNFGISFLLILMLLYSFI 1639
Db 1653 TEEDSGSNGKIKTCBCTKPDSPYPLDGIPTCSSNFGISFLLILMLLYSFI 1704

RESULT 2
Q9NHX1
ID Q9NHX1 PRELIMINARY; PRT; 1694 AA.
AC Q9NHX1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Major merozoite surface antigen.
GN GPI95.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
antigen(gp195)gene of plasmodium falciparum isolate FCCI/HN.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218248; AAF27526.1; -;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
SQ Merozoite.
KW SEQUENCE 1694 AA; 192767 MW; B51634A49E0F6728 CRC64;

Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
Best Local Similarity 61.1%; Pred. No. 1.4e-153;
Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;

Qy 1 MKIIFFLCSLFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQEKMYLNEGTSGT 60
Db 1 MKIIFFLCSLFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQEKMYLNEGTSGT 60

Qy 61 AVTTSTPGSKGSVSGSGSGSVASGSGSVASGSGSVASGSGSNRRTPNSDSSDS 120
Db 61 AVTTSTPGSKGSVSGSGSGSVASGSGSVASGSGSVASGSGSNRRTPNSDSSDS 114

Qy 121 DAKSYADLKHVRVNYLLPTIKELYPQLFDLTNNHMLTCDNIHGFYKLYIDGYEIEINELLYK 180
Db 115 DAKSYADLKHVRVNYLLPTIKELYPQLFDLTNNHMLTCDNIHGFYKLYIDGYEIEINELLYK 174

Qy 181 LNFYDILLRAKLDVNCANDYCOIIPENLKRANELDVLKLVFGYRKPDLNIDKNYGMED 240
Db 175 LNFYDILLRAKLDVNCANDYCOIIPENLKRANELDVLKLVFGYRKPDLNIDKNYGMED 234

Qy 241 YIKNKTITENINELIEESKTKITDNKNATKEEKKLYQAQYDLISYKQLEEAHNLIIS 300
Db 235 YIKNKTITENINELIEESKTKITDNKNADNEEGKKLYQAQYDLISYKQLEEAHNLIIS 294

Qy 301 VLEKRIDTLKKNENIKELDKKINEIK-NPPFANSNGTPTNLLD-KNKKIEEHEKEIKEIA 358
Db 301 VLEKRIDTLKKNENIKELDKKINEIK-NPPFANSNGTPTNLLD-KNKKIEEHEKEIKEIA 358


```
Db 295 VLEKRIIDTLKKNNIKKLEIDIKTDAEKLTTGSKPNPLPENKKKEVEGHEEKIETA 354
Qy 359 KTIKKNIDSLFDPLEYLEYLRKKNKNDISAKVE--FKESTEPN-EYPNGVTPYPLSYND 415
Db 355 KTIKKNIDSLFDPLEYLEYLRKKNKVDVTPKSDPTKSVQIPKVPYNGVIVPLPLTD 414
Qy 416 INNAL---NELNSFGDLINPFDTYKPSKNIYTDN-ERKKEINEIKETIKIEKKI----467
Db 415 IHNSLAADNDKNSYDGLMNP-DTKEKINEKIITDNKERKIFINNIKKOIDLEKKNINHTK 473
Qy 468 ESKKSIEDRSKSLNDITKEYEKLNEIYDSKFNNDIUTNFEKMKGRYSKYVKELTHH 527
Db 474 EONKKLLEDEYERS---KDYEELEKFEYEMFNNDKVDVDFKIFRSARYTYNVEKQKN 529
Qy 528 NTFASVENSKHLEKITKALKYMEYDSLRNIVVEKELKYKNLISKIENEIETLVENIKK 587
Db 530 NFFSSNSVYVQKLLKALSLYEDYSLRGISEKDFNHYITLKTGLEADIAKKLTEEIKS 589
Qy 588 DEEQLFEKK---ITKDKNKPDKILEVSDIVKQVQKVLMMNKIDELKKTQLILKNVELK 644
Db 590 SENKILEKNFKLTHSANAS---LEVSIVKLVQVKVLLKKIEDLRKIELFLKNAQLK 645
Qy 645 HNIHPNSYKQENKQBPYYLIVLKEIDKLKVPMPKVESLINEEKKNIKTEGQSDNSEPS 704
Db 646 DSIHPNIYKPNQKPEPYYLIVLKEVDKLKEFIPKVDMLKKEQAVLSS-----I 696
Qy 705 TEGEITGOATTKPGOAGSALGDSVQAQOAKQAPVPVPVEAKAQVTPPPAPV-- 762
Db 697 TQPLVAASETTEDGHSHTLSQSGETEETETVGHITVITLPTQ---PSPKEVKV 753
Qy 763 -----NNKTENVSKLDYLEKLYEFLNTSYICHKYLVSHTSMNEKILKOYKITBEE 813
Db 754 VENSTEKSNDSNQSALTKTIVLKKDEFLTKSYICHKYLVSNSMDQKLLVYNLTPEE 813
Qy 814 ESKLSKCDPLDLFLFNQNNIPWYSMFDSLNNLSOLFMEIYEKEMVNCVLYKLKNDKTK 873
Db 814 KNELKSCDPLDLFLFNQNNIPAMYSLYDSMNNDLQHLFEFLYQKEMYYLYHLKKEENHK 873
Qy 874 NLEEAQKV-----STSVKTLSSSS-----MQPLSLTPQDQPEVSAND---D 912
Db 874 KLLLEQKQITGTSSTSSPONTVNTAQSATHSNSNQOSNASSTNTQNGVAVSSGPAVVE 933
Qy 913 TSHS-----TNLNSLKLFNILSLGKNKIYQEL-IGQKSSNFENFEKILKDSDTFYNESF 967
Db 934 ESHDPLTVLSISNDLKGIVSLLNGLNKTVPNPLATISTEMEKFYENILKNNDTYFNDOI 993
Qy 968 TNFVSKADDINSLEDSRKKLEEDINKLTKTLQSLPDLNKNYKILKLERLDFDKKTKVCK 1027
Db 994 KQFVNSKSVITGLT-ETQKNALNDEIKKLKDTLQSLPDLNKNYKILKLERLDFDKKTKVCK 1052
Qy 1028 YRMQIKKTLLEQLESKLSLNNPKHVLQNFVSFNNKKEAEIAETENTLENTKILLKH 1087
Db 1053 DKMQIKKTLLEQLESKLSLNNPHVNLQNFVSFNNKKEAEIAETENTLENTKILLKH 1112
Qy 1088 YKGLVKYNGESSPLKTLSEESIQTEDNTVASLENKVLKLECKLKDNLNLEKKKLSYLUS 1147
Db 1113 YKGLVKYNGESSPLKTLSEVSIQTEDNTVANLEKFRVLSKDGLKNDNLHLGKKLSFLS 1172
Qy 1148 SGLHLHIAELKVIKKNKNTGNSPENNNDVNALESYKKLPECTDVATVYSE-----1201
Db 1173 SGLHLHITELKEVIRKKNKNTGNSPENNKKVNEALKSYENFLPE-AKVTVTVTPQPDVT 1231
Qy 1202 -----SGSDTLQSQPKKPASTHVGAESNTITTSQWDDVDDVIVPIFGSEEDY 1253
Db 1232 PSPLSVRVSSSGSTKEETQITSGSLLTLEQVQSQNYDEEDSLVLPVIFGESEND 1291
Qy 1254 DDLGQVVTGEAVTPSVIDNLISKIENEYEVLYKPLAGVYRSLKQLENNVMTFNWVKD 1313
Db 1292 EYLDQVVTGEAISVT-MDNILSGFENEYDVIYKPLAGVYRSLKQIEKNIITFNLNLD 1350
Qy 1314 ILSNRFNKENFNKVLSDLPYKDLTSSNYYVKQPYKFLNKEKRDCKFLSSNYIKDSTD 1373
Db 1351 ILSNRLKRRKRYFLDVLSDLMQFKHSSNEYIIEDSFLLNSEQNTLLKSYIKESVE 1410
```

```
Qy 1374 TDINEANDVLGYKILSEKYSKSDLSIRKYY-----NDKQGENEY 1414
Db 1411 NDIKFAQEGISYEVKVLAKYKDDLESIKKVIKEEKEFFSSPPTTPPSPAKTDEQKESF 1470
Qy 1415 LPLFNNIETLYKTVDNKIDLFIHLEAKVLYNTYKSNVEYKIKELNYLKTIODKLADPK 1474
Db 1471 LPLFTNIETLYNVLNKIDDIYLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFPK 1530
Qy 1475 KNNFVGADLTSTDYNNHNLTKFLSTGMVFENLAKTVLSNLDGNLQGLMNLISQHCYK 1534
Db 1531 NTNDFEATKLLNDOTKMDLKGKLLSTGLV-QNFPNTIISKLEIGKFDQMLNLISQHCYK 1589
Qy 1535 KQCPNSGCFRHLDRERECKCLLNKYQEGDKCVENPNTCENNNGCCDADAKTEEDSGS 1594
Db 1590 KQCPNSGCFRHLDRERECKCLLNKYQEGDKCVENPNTCENNNGCCDADAKTEEDSGS 1649
Qy 1595 NGKKTCECTKPDSPVPLFDGIFCSCSNFLGIFSLILMLILYSFI 1639
Db 1650 NGKKTCECTKPDSPVPLFDGIFCSCSNFLGIFSLILMLILYSFI 1694
```

RESULT 3

```
Q9TZT5
ID Q9TZT5 PRELIMINARY; PRT; 1694 AA.
AC Q9TZT5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Merozoite surface protein 1.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HN1;
RA Jiang G., Liu R., Daubenberger C.A., Pluschke G.;
RT "Sequence analysis of the MSP 1 gene of Plasmodium falciparum from
RL Hainan, China."
RL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung 17:294-297(1999).
DR EMBL; AF062348; AAC72884.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
SQ SEQUENCE 1694 AA; 192795 MW; 84CFC0E709F5673B CRC64;
```

Query Match 61.3%; Score 5162.5; DB 5; Length 1694;
Best Local Similarity 61.1%; Pred No. 1.4e-153;
Matches 1054; Conservative 226; Mismatches 328; Indels 117; Gaps 27;

```
Qy 1 MKIIFFLCSFLFIINTQCVTHESYQELVKLEALEDAVLTYSLFQKEKMLNEGTSQT 60
Db 1 MKIIFFLCSFLFIINTQCVTHESYQELVKLEALEDAVLTYSLFQKEKMLNEGTSQT 60
Qy 61 AVTTSTPGSKGVSAGSGSGSVASGSGSVASGSGSVASGSGSGSNRRTPNSDSSS 120
Db 61 AVTTSTPGSGSVTSGGSGSVA---SVASGG---SGGSVASGSGSGSNRRTPNSDSSS 114
Qy 121 DAKSVADLKHRVNYLTTIKELKYPOLFDLTNHLMLTCLDNIHGFYKYLIDGYEINELLYK 180
Db 115 DAKSVADLKHRVQNYLFTIKELKYPELFDLTNHLMLTCLDNIHGFYKYLIDGYEINELLYK 174
Qy 181 LNFYFDLLRAKLVNDVANDYCOIPFNKLRANELDVLKLVFGYRKRPDLNIDNVGKMD 240
Db 175 LNFYDILLRAKLVNDACANSYCOIPFNKLRANELDVLKLVFGYRKRPDLNIDNVGKMD 234
Qy 241 YTKKNNKTITENINELIESKKTIDKNNKNTKBEKKKLYQAQYDLSIYNKQLEEAHLIS 300
Db 235 YTKKNNKTITANINELIEGSKTIDQKNADNEEGKKLYQAQYNLFYNNKQLEAHLIS 294
Qy 301 VLEKRIIDTLKKNNIKKLEIDIKTDAEKLTTGSKPNPLPENKKKEVEGHEEKIETA 358
Db 295 VLEKRIIDTLKKNNIKKLEIDIKTDAEKLTTGSKPNPLPENKKKEVEGHEEKIETA 354
```



```
Qy 359 KTIKFNIDSLFTDPLEYLLREKKNIDISAKVE--TKESTEPN-EYPNGVTYPLSYND 415
      |||||
Db 355 KTIKFNIDSLFTDPLEYLLREKKNYDVTPKSDPTKSVQIPKVPYPNGIVPLPLTD 414
Qy 416 INNAL---NELNSFDGLNPDPYTKPEKSNITDN-ERKKFNETKEKIKIEKKI----- 467
      |||||
Db 415 IHNLAADNRKNSYGLMNP-DTKEINEKIITDNKERKIFINNIKKQIDLEEKINHTK 473
Qy 468 ESKKSYEDRSKSLNDITKEYEKLEINEYDSKFNNDILTFEKMGMGRYSYKVEKLTH 527
      |||||
Db 474 EONKLLLEDYKES----KKOFEELKEFEMKFNNDKVDVDFKISARYTINVEKQYN 529
Qy 528 NTFASYENSKNLEKLTALKYMEDYSLRNIVVEKELYSYKNLISKIENEIETLVENIKK 587
      |||||
Db 530 NKFSNNNSVNVOKLKALSYLEDYSLRKGISSEKDFNHYTTLTGLEADIKKLTTEEKS 589
Qy 588 DEBQLFERK---ITKDEKNKPEKILEVSDIVKVQVQVLLNNKIDELKKTOLILKNVELK 644
      |||||
Db 590 SENKILEKNFRGLTHSANAS----LEVSDIVKLOVQVLLKIBDLRKIELFLKNAOLK 645
Qy 645 HNIHVPNSYKOEKQEPYLLIVLKEIDKLVEMPKVESLINEEKKNIKTBEQSDNSEPS 704
      |||||
Db 646 DSIHVPNIYKQNKPEPYLLIVLKEVDKLEFIPKVKMDLKKQOAVLSS-----I 696
Qy 705 TEGITGQATKPGQAGSALGDSVQAQAEQKQAOQPPVPVPPEAKAQVPTPPAPV-- 762
      |||||
Db 697 TQPLVAASETTEDGGHSTHTLSQSGETEETETVGHVTTVITLPTQ---PSPKPEKV 753
Qy 763 -----NKNTEVSKLDYLEKLYEPLNTSYICHKYILVSHSPMKNKILKQYKITEE 813
      |||||
Db 754 VENSTEHKSNDSQALTTVYLLKLDLFTKSYICHKYILVSNSSMDQKILLEVNLTP 813
Qy 814 ESKLSSCDPLDLFNIONNIIPWYSMFDLSNLSQLEPMEIYEXEWCNLYKLKDNDAIK 873
      |||||
Db 814 KNEKSCDPLDLFNIONNIIPWYSMFDLSNLSQLEPMEIYEXEWCNLYKLKDNDAIK 873
Qy 874 NLLEAKKV-----STSVKTLSSS-----MQPLSLTPQDKPEVSAND---D 912
      |||||
Db 874 KLEEQKQITQTSSTSSPGNTTNTAQSAHNSQONQSNASSINTQNGVAVSSGPVVE 933
Qy 913 TSHS-----TNLNSKLPELNLISGKNKIYOEL-IGOKSSENFYEKILKDSDTFYNESF 967
      |||||
Db 934 ESHDPLTVLSISNDLKGTVSLNLGNKTKVPNPLTISTEMEKFEYENILKNNDTYFNDI 993
Qy 968 TNEVSKADDIINLNDSEKRLKLEEDINKLAKTLQSLDLYNKYKLLERLFDKKTGYK 1027
      |||||
Db 994 KQFVKSNSKVITGLT-ETQKALNDEIKKLDTLQSLFDLYNKYKLLERLFDKKTGYK 1052
Qy 1028 YKMQIKKLTLLKEQLESKLSLNNPKHVLQNFVSFFFNKKKBAEIAETENTLTKILLKH 1087
      |||||
Db 1053 DKMQIKKLTLLKEQLESKLSLNNPHVNFVSFFFNKKKBAEIAETENTLTKILLKH 1112
Qy 1088 YKGLVYKNGSSPLKTLSESIOTEDNYASLENFKVLKLEGLKDNLNLEKKKLSVLS 1147
      |||||
Db 1113 YKGLVYKNGSSPLKTLSEYSIQEDNYANLEKFRVLKIDGKLNDNLHLGKKKLSPLS 1172
Qy 1148 SGLHLLTAELEKVIKKNKNTGNSPENNTDYNNALESYKFKLPCTDVATVUSE----- 1201
      |||||
Db 1173 SGLHLLTAELEKVIKKNKNTGNSPENNTDYNNALESYKFKLPCTDVATVUSE----- 1231
Qy 1202 -----SGSDTLLEQSQPKPASTHGAESNTITTSQVDDVEDVDVITVPIFGESEEDY 1253
      |||||
Db 1232 PSLSVRVSGSGGSKETQIPTSGSLTLELQVQVQSONYDEEDSLVVLPIFGESEEND 1291
Qy 1254 DDLQGVVTGEAVTSPVIDNLSKTIENEYEVLYLKLPLAGVYRSLLKQLENNVMTFNVNKD 1313
      |||||
Db 1292 EYLDQVVTGEAISVT-MDNILSGFENEYDVYLYLKLPLAGVYRSLLKQLEKNIITENLND 1350
Qy 1314 ILNSRFNKNREKFNKLVESDLIPYKDLTSSNVVVDKPYKFLNKEKRDKFLSSYNIKSID 1373
      |||||
Db 1351 ILNSRLKRRKYFLVDLESMDLQFHHISSNEYIIEFSDSKLLNSEQKNTLLSKYKIKESVE 1410
```

```
Qy 1374 TDINFANDVLGYKILSEKYSKSDLSIRKYYI-----NDKQGENEKY 1414
      |||||
Db 1411 NDIKFAEGISYSEKVLAKYKDDLESIKKVIKEEKEFPSSPTTPPSPAKTDEQKESKF 1470
Qy 1415 LPFLNNIETLYKTVNDKIDLFVHLEAKVNLTYEKSNNVEVKIKELNYLKTQDKLADPK 1474
      |||||
Db 1471 LPFLNIIETLYNNLVNKIDDLINLAKINDCNVEKDEAHVKITKLSDLKAIDDLK 1530
Qy 1475 KNNPVGADLSTDYNNHNNLLTKFLSTGMVFNENLAKTVLSNLLDGNLQGLNLSOHCVK 1534
      |||||
Db 1531 NTNDFEATKLLINDOTKMDLGLKLLSTGLV-QNFPTTIISKLEIEGKFOMLNLSHQCVK 1589
Qy 1535 KQCPNSGCFRHLDEREBECKLLNKKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGS 1594
      |||||
Db 1590 KQCPNSGCFRHLDEREBECKLLNKKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGS 1649
Qy 1595 NGKKTICBCTPDPSPYPLFDGIFCSCSNFLGISTFLLILMLILYSFI 1639
      |||||
Db 1650 NGKKTICBCTPDPSPYPLFDGIFCSCSNFLGISTFLLILMLILYSFI 1694
RESULT 4
Q25922 PRELIMINARY; PRT; 1720 AA.
AC Q25922;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Precursor of the major merozoite surface antigens.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NC NCBI_TaxID=5843;
RN [1]
SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R., Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=NF54;
MEDLINE=96123395; PubMed=8577332;
RA Pan W., Tolle R., Bujard H.;
RT "A direct and rapid sequencing strategy for the Plasmodium falciparum
  antigen gene gp190/MSAL.";
DR MBL: Z35327; CA84556.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW Merozoite; Signal.
FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;
```

Query Match 60.6%; Score 5107.5; DB 5; Length 1720;
Best Local Similarity 60.2%; Pred. No. 7.3e-152;
Matches 1052; Conservative 216; Mismatches 343; Indels 137; Gaps 29;

```
Qy 1 MKIIFFLCSFLFFINTOCVTHESYQELVKKLEALEDVLTGYSLFQEKWVLE---GT 57
      |||||
Db 1 MKIIFFLCSFLFFINTOCVTHESYQELVKKLEALEDVLTGYSLFQEKWVLE---GT 60
Qy 58 SGTAVTTSTPGSKGSVAGSGGSGSVAGSGSVAGSGSVAGSGSGN---SRRTN--- 112
      |||||
Db 61 KGASQAQSGASQSG--ASQAQSGASQAQSGASQAQSGTSGPSGSGTSPSSRSNTLP 118
Qy 113 -----PSDNSDSDAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMTLCDNTHG 163
      |||||
Db 119 RSNSTSGASPPADASDSDAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMTLCDNTHG 178
```



```
QY 164 FYXLDGYEINELYLKLFYFDLRAKLVNDVANDYCOIPNLKIRANELDVLKLVFG 223
Db 179 FYXLDGYEINELYLKLFYFDLRAKLVNDVANDYCOIPNLKIRANELDVLKLVFG 238
QY 224 YRPDLNIDNVGKMDYTKKNNKTITENIELIESKKTIDKNNKATKEEEKKLYQAOY 283
Db 239 YRPDLNIDNVGKMDYTKKNNKTITIANIELIEGSKTIDQNNKADNEEGKKLYQAOY 298
QY 284 DLSYKQLEAEHNLISVLEKRIDTLKKNENIKELDKINEIKNPPANSGTPTNLLDK 343
Db 299 DLSYKQLEAEHNLISVLEKRIDTLKKNENIKELDKINEIKNPPANSGTPTNLLDK 358
QY 344 NKKIEHEKEIEIAKTIFENIDSFTDPLEYILREKKNIDISAKVE--TKESTEPN 401
Db 359 NKKIEHEKEIEIAKTIFENIDSFTDPLEYILREKKNIDISAKVE--TKESTEPN 418
QY 402 EYPNGVTPVLSNDINNAL---NELNSFGDLINPDYTKEP-SKNIYTDN-ERKKFINE 455
Db 419 VYPNGIVVPLPTDTHNSLAADNKNYSYGLMNP--HTKEKINEKIITDNKERKIFINN 476
QY 456 IKKIDLEKKNHNTKEQNKLLDEYKS---KKDYBELLEKFEYEMKFNNDVVDK 511
Db 477 IKKIDLEKKNHNTKEQNKLLDEYKS---KKDYBELLEKFEYEMKFNNDVVDK 532
QY 512 MMGKRYSVKVELTHHTTASVENSKHNLKLETKALKYMEDYSLRNIVVVEKELKYKNLI 571
Db 533 IFSARYTVNEKQRYNKFSSNNSSVYVQKLKALSLEYDSLKRGISEKDFNHYITLK 592
QY 572 SKTENIEITLVENIKKDEQLEK---ITKDNKPDKEILEVSDIVKVOVKVLLMKNKI 628
Db 593 TGLEADIKKLEIKESSEKNIIEKNPKGLTHSANGS----LEVSDIVKVOVKVLLMKNKI 648
QY 629 DELKKTQLILKVNELKHNTVHVPNSYKQENKQBPYYLIVLKEIDKLKVPMPKVESLINEE 688
Db 649 EDLRKLELFLKNAQLKDSIHVPNIYKPNQKPEYYLIVLKEIDKLKVPMPKVESLINEE 708
QY 689 KNIKTGOSDNSEPTGEITGOATTKGQOAGSALGDSVQAOAQEQKQAO-----P 742
Db 709 QAVLSS-----ITQPLVAASETTEDGHSHTHTLSQGETEVTEETEETVGHPT 759
QY 743 PVPVPEAKAQPPTPAVP-----NNKTENVSKLDYLEKLYEFLNTSYICHKY 791
Db 760 TWITITLPPQ---PSPPEVKVYVENSIEHKSNDNSQALTKYVLKKLDELTKSYICHKY 816
QY 792 ILVSHSTMKEILKOYKTKIEESKLSKCDPLDLFNQNNIPVYMSFDSLNLSLSQLF 851
Db 817 ILVSNSSMDQKLEVLNTPPEENELKSCDPLDLFNQNNIPVYMSFDSLNLSLSQLF 876
QY 852 MEIYKEMVNCYLKLDNDKIKNLLEAKV-----STSVKTLSSSS----- 893
Db 877 FELYQEMTYLHLKKEENHKKLEEQKQITGTSSTSSPGNTTNTAQSATHSNQOQ 936
QY 894 MOPLSLTPDQKPEVSAND---DTSHS---TNLNSLKLFEINLSLGKKNYQOEL- 945
Db 937 SNASNTQNGAVVSGPAAVVEESHDPVLVLSINDLKGIVSLNNGKNTKVPNPJTIST 996
QY 946 KSENFYKILKDSDFYFNESETNFVKSADDDINSLNDESRRKLEEDINKLTKTLQLSF 1005
Db 997 TEMEKYEYENLKNNDYFNDDIKQFVKSNSKYITGLT-ETQKNALNDEIKLKTDLQLSF 1055
QY 1006 DLYNKYKLERLFDKKTGVGYKMQIKKLTLLKEOLSKLSLNNPKHVLFQNFVFPNK 1065
Db 1056 DLYNKYKLERLFDKKTGVGYKMQIKKLTLLKEOLSKLSLNNPKHVLFQNFVFPNK 1115
QY 1066 KKEAEIAETENTLENPKILLKHYKGLVXYNGESSPLKTLSESIQTEDNYASLENFKVL 1125
Db 1116 KKEAEIAETENTLENPKILLKHYKGLVXYNGESSPLKTLSESIQTEDNYANLEKFRVL 1175
QY 1126 SKLEGKLDNLNLEKKSLSLSSGLHLIAELKEVTKNNKNTGNSPENNTOVNALESY 1185
Db 1176 SKIDGLNDNLNLEKKSLSLSSGLHLIAELKEVTKNNKNTGNSPENNTOVNALESY 1235
QY 1186 KKFLEPGTOAVTVVSE-----SGSDTLQEQSQPKKPASTHVGAESNTITTSQ 1231
```

```
Db 1236 ENFLPE-AKVTVTVTPQPDVTPSPLSVRVSGSGSTKEETQIPTSGSLTLEQQVVQLO 1294
QY 1232 NVYDEVDVVIIPIFGESEEDVDLGOVVTGEAVTPSVINDILSKITENEYEVLYKLPLAG 1291
Db 1295 NYDEEDSVLPIFGESEEDNDDEYLDQVVTGEAISVT-MDNILSGFENYDVIYLKPLAG 1353
QY 1292 VYRSKQLENNVMFNVKDIILNSRFNKNFKNVLESDLIPYKDLTSSNYVVVADPYK 1351
Db 1354 VYRSKQLENNVMFNVKDIILNSRFNKNFKNVLESDLIPYKDLTSSNYVVVADPYK 1413
QY 1352 FLNKKRDLFSLSYNIKOSIDTDINFANDVLYGKILSEKYSKLSLDLSKKYI----- 1404
Db 1414 LLNSEQNTLLSKYIKESVENDIKFAQEGISYSEKVLAKYKDDLESIKVKEKEKF 1473
QY 1405 -----NDKOGENEKYLPLNNIETLYKTVDNKIDLVFVHLAKVLNLYEKS 1451
Db 1474 PSSPPTPPSPAKTDEQKESKFLPLTNIETLYNNLVNKKIDYLLINLAKKINDCNVEKD 1533
QY 1452 NYEVAIKELNYLKTODKLADFKNNNFYGIADLSTDYNNHNLTKFLSTGMVFENLAKT 1511
Db 1534 EAHVITKLSDLKAIDDKIDLFKNPYDPEAIKKLINDTKDMLGKLLSTGLV-QNFPNT 1592
QY 1512 VLSNLDGNLQMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPN 1571
Db 1593 IISKIEGKFDMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQEGDKCVENPN 1652
QY 1572 PTCNENNGCCDADAKTEEDSGNKKITCECTKPDSPYPLFDGIFGSSNFGLISPLLL 1631
Db 1653 PTCNENNGCCDADAKTEEDSGNKKITCECTKPDSPYPLFDGIFGSSNFGLISPLLL 1712
QY 1632 MLILYSFI 1639
Db 1713 MLILYSFI 1720

RESULT 5
Q8T6A9 PRELIMINARY; PRT; 656 AA.
AC Q8T6A9,
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCB-1;
RA Li X.R., Liu S.C., Chishti A.H., Oh S.S.;
RT "83 kDa subfragment of MSP-1 gene of Plasmodium falciparum isolate FCB-1.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF480451; AAL84455.1;
FT NON_TER 1
FT NON_TER 656
SQ SEQUENCE 656 AA; 75698 MW; E1EF44ACE341FD0C9 CRC64;

Query Match 39.9%; Score 3361; DB 5; Length 656;
Best Local Similarity 100.0%; Pred. No. 7.6e-98;
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 VTHESYQELVKLEALEDAVLGTGYSLFQEKVYLVNGETSGTAVTTSTPGSKGVSASGGSG 79
Db 1 VTHESYQELVKLEALEDAVLGTGYSLFQEKVYLVNGETSGTAVTTSTPGSKGVSASGGSG 60
QY 80 GSVASGSGSVASGSGSVASGSGSGNSRRTPNSDSSDAKSYADLKHRYNRYLLTI 139
Db 61 GSVASGSGSVASGSGSVASGSGSGNSRRTPNSDSSDAKSYADLKHRYNRYLLTI 120
QY 140 KELYPQLEFDLNNHMLTCLDNIHGFKYLDIGYEENELLYKLNIFYDILLRAKLVNDV 199
```



```
Db 121 KELYKPOLFDLTHMLTLCNDNHGKYLIDGYEINELLYKLFYDILLRAKLDVNCAND 180
QY 200 YCOIPFNKIRANELDVLKLVGKRPDLNDIKDNVGMEDYIKNNKKTNIENIELIES 259
Db 181 YCOIPFNKIRANELDVLKLVGKRPDLNDIKDNVGMEDYIKNNKKTNIENIELIES 240
QY 260 KKTIDKNKATKEEKKKLYQAYDLSYNKOLEEAHNLISVLEKRIDPLKKNENIKELL 319
Db 241 KKTIDKNKATKEEKKKLYQAYDLSYNKOLEEAHNLISVLEKRIDPLKKNENIKELL 300
QY 320 DKINEIKNPPANGNTPTNLDKNKKIEEHEKEIKEIAKTIKFNIDSFTDPLEEYVL 379
Db 301 DKINEIKNPPANGNTPTNLDKNKKIEEHEKEIKEIAKTIKFNIDSFTDPLEEYVL 360
QY 380 REKNKIDISAKVETKESTPEPNEYNGVTPUSYNDINNALNELSFGDLINFPDYTKEP 439
Db 361 REKNKIDISAKVETKESTPEPNEYNGVTPUSYNDINNALNELSFGDLINFPDYTKEP 420
QY 440 SKNIYTDNERKKFINEIKIEKKIESDKKSYEDRSKSLNDITKEYEKLLEIYDSK 499
Db 421 SKNIYTDNERKKFINEIKIEKKIESDKKSYEDRSKSLNDITKEYEKLLEIYDSK 480
QY 500 FNNIDLNPFKMMGKRYKYVEKLTHTNTFASYSKHNLEKLTALKYMEDYSLRNIV 559
Db 481 FNNIDLNPFKMMGKRYKYVEKLTHTNTFASYSKHNLEKLTALKYMEDYSLRNIV 540
QY 560 VEKELKYKNLSKIENETELVENIKKDEOLFPEKLTIDENKPEKLTILEVSDIVKQV 619
Db 541 VEKELKYKNLSKIENETELVENIKKDEOLFPEKLTIDENKPEKLTILEVSDIVKQV 600
QY 620 QKVLLMKNIDELKKTQLILKNVELKHNHVPNSYKQENKQEPYIIVLKKIEIDKLK 675
Db 601 QKVLLMKNIDELKKTQLILKNVELKHNHVPNSYKQENKQEPYIIVLKKIEIDKLK 656

RESULT 6
ID Q26194 PRELIMINARY; PRT; 1751 AA.
AC Q26194;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE P200 protein precursor.
GN P200.
OS Plasmodium vivax.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5855;
RW [1]
SEQUENCE FROM N.A.
RC STRAIN-SAL-1;
RX MEDLINE=92158013; PubMed=1371329;
RA Gibson H.L., Tucker J.E., Kaslow D.C., Kretzli A., Collins W.E.,
RA Kiefer M.C., Bathurst I.C., Barr P.J.;
RT "Structure and expression of the gene for P200, a major blood-stage
RT surface antigen of plasmodium vivax."
RL Mol. Biochem. Parasitol. 50:325-334(1992).
DR EMBL; M75674; AAA29735.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Signal.
FT SIGNAL.
FT CHAIN
SQ SEQUENCE 1751 AA; 196120 MW; 4162F9883F29D8A6 CRC64;
```

```
Query Match 36.7%; Score 3093.5; DB 5; Length 1751;
Best Local Similarity 38.2%; Pred. No. 4.4e-89;
Matches 704; Conservative 316; Mismatches 517; Indels 307; Gaps 41;

QY 1 MKIIFLCSFLFIINTOCVTHESYQELVKLEALEDAVLTCYSLFOKEKMYLNEGTSCT 60
Db 1 MKALLFLSFIFVTKOCET-ESTKQLVAKLDKLEALVVDGTYELFHKKL----- 50
```

```
QY 61 AVTTSTPGSKGVSAGSGSGSVASGSGSVASGSGSVASGSGSRRTPNSSDS 120
Db 51 -----GENDIKVETNASANNNN 68
QY 121 DAKSYADLUKHRVRYNLLTIKELKYPQLFDLTNHLMLTLCNDNIHGFKYLDIGYEINELLYK 180
Db 69 NOVSV--LTSKIRNLSKFLQIPGHTDOLLHILRELAVEPNGIKYLVSEYEFNOLMHV 126
QY 181 LNFYDILLRAKLDVNCANDYCOIPFNKIRANELDVLKLVGKRPDLNDIKDNVGMED 240
Db 127 INFHYDILLRAKLDVNCANDYCOIPFNKIRANELDVLKLVGKRPDLNDIKDNVGMED 186
QY 241 YTKKKNKTITENEL-IEESKK----- 261
Db 187 FITKKNKIITIKNISDLIIAENKKRSGHPTTTNAGCTQPANGSIAASSETTQISGSSNSG 246
QY 262 -----TIDKKNKATKEEKKKLYQAYDLSYNKQ 291
Db 247 SSSTGSSNSGSSSTGSGTGTGQSPPAADASTNANYEAKKIYQAVNTIFTYTNQ 306
QY 292 LEEAHNLISVLEKRIDPLKKNENIKELDKINEIKNPPA---NSGNTPTNLDKNKKIE 348
Db 307 LQEAQKLAVLEKRVKVLKEHKDIKVLLEQVAKKEKLPDYPNTNLTNVHKEAESKIA 366
QY 349 EHEKEIKEIAKTIKFNIDSFTDPLEEYVLREKKNKI-DISAKVETKESTPEPNE----- 402
Db 367 ELEKKIEIAKTIKFNIDSFTDPLEEYVLREKKNKI----- 426
QY 403 ---YNGVTPUSYNDINNALNELS---FGDLINFPDYTKEPSKI-YTDNERKKFINE 455
Db 427 KETYPHGYSYLAENSIEYELIEKIGSDETFDGLQNP-DGQKQPKKGLINETKRELBLEK 485
QY 456 IKEKIKIEKKIESDKKSYEDRSKSLNDITKEYEKLLEIYDSKFNNNIDLTNFEKMMCK 515
Db 486 IMNKKIEBKLPNKKYEYKLYVEAKVNEFKPAPNFHYEARLONTLIVENKFDFFK 545
QY 516 RYSYKVEKLTHTNTFASYSKHNLEKLTALKYMEDYSLRNIVVEKELKYKNLSKIE 575
Db 546 REAYMEKKKLES--CSYEQNSNLINKKQLTYLEDYVLRDIADDEIKHESFMEMWKLK 603
QY 576 NEIETLVENIKKDEOLFPEKLTIDENKPEKLTILEVSDIVKQVQKVLMMKNIDELKKTQ 635
Db 604 SEIYDLAQEIRKNNK-----TIENK-----FDPSGVVELOVQKVLIIKIEALKNYQ 652
QY 636 LILKNVELKHNHVPNSYKQENKQEPYIIVLKKIEIDKLKLVEMPKVESLINEBEK----- 690
Db 653 NLLKNKVKDDLYIPKYKTSKPEPYIIVLKKIEIDKLKDFIPKIESMIATEKKNPTVA 712
QY 691 ---NIKTEGOS--DNSEPSTEG-----EITQATTKPG-----QAQSA-LE 726
Db 713 AADIVAKGOSLRGASETGTTGNTVNAQTAVQVQHQVNVAVTVQPGTTGHOAQGEAETQ 772
QY 727 GDSVQAQAOEQKQA-----OPPVPVPV-PE-AKAQVTPPPAPVNNKTENVSK 771
Db 773 TNSVQAQVQQTTPAGAGGOVASTQISOAPATQSPAPAPAPSTPAAAVAPAPTMSK 832
QY 772 LDYLEKLYEFLNTSYCHKYILVSHSTNMKEIKLYKITKEESKL--SSCDPLDLLLLNI 829
Db 833 LEYLEKLLDFLKASAYACHKHIFVNSTMMKELLDQYKLNADSEONKINETKCDLDELFFV 892
QY 830 QNNIPWYSMFDLSNLSQLFMEIYKEMWCNLYKLKND-KIKNLLSEAK-----VS 883
Db 893 QNNLPAMYIYDSMSNLQNLIELYQKEMVNIYKNTDKKKIAFLAFLEKSKAAAPAQ 952
QY 884 TSVKTLSSSSMOPLSIT-PQDKPEVSANDDTSHSTNLSNLSKILFENILSLGKNKIYQEL 942
Db 953 SAAKPSGAGTTPVTPTAPVTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1010
QY 943 IGKKSSENFYEKILKSDTFYNESTFNFKKADDTINSINLSDESKRKKLEEDINKLKTITQ 1002
Db 1011 ADTAQVEFYKHLSQLDK-YNDYQKFLQESQDEITKM-DETKWKALGAEIEELKKLQ 1068
QY 1003 LSFIDLNYKYLKLERLFDKKKTGVGYKMOIKKLTLLKEQLESKLSNLPKHVLFQNSVF 1062
```



```

1069 VSLDHYGKYLKRLERLLKKNKNSKQIKLTSLNKLERQNLNNPTSLKKNYTA 1128
1063 FNKKKAETAEENTLTKLLKHVKGLVYNGSSPLKTLSESIOTENVASLENF 1122
1129 FNKRETERKEVENTLKNTEILLKYKARAKYIIGPFPPLKTLSESMOKEDNYLNEKF 1188
1123 KVLKLEGLKLDNLNLEKKLSYSSGLHHLIAELKEVKIKNNYTNPSSENNTDYNAL 1182
1189 RVLRLGLRGLKGLNLEKENISYSSGLHHLVLTLEKILKKNKYSNDHTKNAKVEAL 1248
1183 ESKYKLPEDGTAVTVVSSGSDTLEQS-----QPKKPA- 1216
1249 QAYOEILPK-----VTFQEGASTTAATLPVTVPSAVPGGLPGAGVGAAGLTPPPAG 1302
1217 ---STHVA-----ESNTITTSQNVDEVDVLIIPFGESEDEYDGLGVVTVGEA--- 1264
1303 SVPATGGAAGSTENVAAKADYAEYDKVIALFLFGNNDODGEE-DQVTTGEASEA 1361
1265 ---VTPSVIDNLSIKENIEYEVLYLKLPLAGVYRSKQLQENNVMTFNVNKOILNSRFNK 1321
1362 PEILVPAGI-----SDYDVYLKPLAGMYKTIKKQLENHVNAFNTNITDMLDSRLKK 1413
1322 RENFKNVLSDLIPYKDLTSSNVVVDYPKFLNKRDRKFLSSYNIKSIDTIDINFAND 1381
1414 RNYFLEVLSNLDNPPFYSSGEYIIRDPYKLLDLEKKKLLIGSYKYIGASIDMDLATAND 1473
1382 VLGYYKILSEKYSKSLDLSIK---KYIND-----KQEN 1411
1474 GVTYNKMGLYKTHLDGVKTEKKYVEDDIKKQDEBKLLGNVNSQDSKNEPIAKKAEL 1533
1412 EKYLPFLNNIETLYKTVDKIDLVILHLEAKVLYNYEYSNVEVKIKELNYLKTIODKLA 1471
1534 EKYLPFLNSLQKEYESLVSKVNTYTDNLKVKVINNCKLEKEAETVYKQLQYKNKMDKLE 1593
1472 DPKNNNFVGIADLSTDYNNHNLTLFELSTGMVFENLAKTVLSNLQGNLQGNLTSQHQ 1531
1594 EYKK-----SBKKNVSSGGLLEKLMKSKLIKENESEILSQLLNVQTLTLTMSSEHT 1646
1532 CVKQKOPNSGCPRHLDRECECKLLNLYKQEGDKGVENPNTCENNNGGCDADAKCTEED 1591
1647 CIDTVNPDNAACVRYLDGTEHRCLTLFREEGKGVPSASVYTKDNNGGCAPEACKMTD 1706
1592 SGSGNKKITCECTKPDSPYPLDGFICSSNFLGISFLILMLIL 1635
1707 S-----NKIVCKCTKEGSEPLFEGVFCSSSFLSLSLFLMLLFL 1746

RESULT 7
Q25961 ID Q25961 PRELIMINARY; PRT: 1087 AA.
AC Q25961;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major surface antigen (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71;
RX MEDLINE=92275047; PubMed=1592091;
RA Olafsson P., Matile H., Certa U.;
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RT RO-71 isolate is recognized by mouse antibody against the
RT nonrepetitive repeat block of RO-33.";
RL Exp. Parasitol. 74:381-389(1992).
DR EMBL; X61930; CAA43932.1; -.
DR InterPro; IPR000087; Collagen.
FT NON_TER 1087 1087
SQ SEQUENCE 1087 AA; 123911 MW; 5518852133C01B33 CRC64;
```

```

Query Match 35.7%; Score 3003.5; DB 5; Length 1087;
Best Local Similarity 57.4%; Pred. No. 1.8e-86;
Matches 638; Conservative 147; Mismatches 238; Indels 89; Gaps 22;

QY 1 MKTIIFLCFLFIINTQCVTHESYQELVKKLEALEDAVLTYGSLFQKEKMWLINE----- 55
DB 1 MKTIIFLCFLFIINTQCVTHESYQELVKKLEALEDAVLTYGSLFQKEKMWLINEBEITF 60
QY 56 -----GTSGTAVTSTPGSGKSVASGGSGSVASGGSGSVASGGSGSVASGGSGS 108
DB 61 KGASAGSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGT 118
QY 109 RRTNPSDSDSDAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMTLCDNTHGPKYLI 168
DB 119 GAIPAD-ASDSDAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMTLCDNTHGPKYLI 177
QY 169 DGYEENELLYKLNFYFDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPL 228
DB 178 DGYEENELLYKLNFYFDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPL 237
QY 229 DNKDNVGMKEDYIKKNKKTININELIBESKKTIDKNKNATKEEKKKLYQAYQDLSIY 288
DB 238 DNKDNVGMKEDYIKKNKKTIANINELIBESKKTIDKNKNADNEBCKKKLYQAYQDLSIY 297
QY 289 NKOLEBAHNLISVLEKRIDLTKKNENIKELLDKINEIK-NPPPSANGTNPNTLLDNKKKI 347
DB 298 NKOLEBAHNLISVLEKRIDLTKKNENIKELLDKINEIK-NPPPSANGTNPNTLLDNKKKI 357
QY 348 EBHEKEIKELIAKTIKENIDSLFTDPLEYLLREKKNIDISAKVE--TKESTEPN-EYP 404
DB 358 EBHEKEIKELIAKTIKENIDSLFTDPLEYLLREKKNIDISAKVE--TKESTEPN-EYP 417
QY 405 NGVTYPLSYNDINNAL---NELNSFGDLINPFDYTKPSKNIYTDN-ERKKFINEIKEKI 460
DB 418 NGIYPLPLDTHNSLAANDKNYSGLMNP-DTKKINEKIITDNKERKFINNKKQI 476
QY 461 KIEKKI-----ESDKSKYEDRSKSLNDITKEYEKLNEIYDSKFNNDITNFEKMMGR 516
DB 477 DLEEKINHTKEQNKKLLDEYKS-----KKOYEELLEKEFYENKFNNDVVDKIFSR 532
QY 517 YSYKVEKLTHHTTFASYENSKHNLEKLTAKALYMEDYSLRNIVVEKELYKKNLSKIEIN 576
DB 533 YTYNVEKQIYNNKFSNNSVNVQKKALSYLEDSLRKGISEKDFHNYTLTKGLEA 592
QY 577 ETEVLNENIKDEQLEFEKK---ITKDENKPOEKILEVSDIVKQVQKVLNMLIDELKK 633
DB 593 DIKKLTEEIKSSENKILEKNFKLTHSANAS-----LEVSDIVKLVQKVLNKKIEDLRK 648
QY 634 TQILKNVLEKHNIHVPSYKQENKOEYLYLVKKEIDKLVFMPKVESLNEEKKNIK 693
DB 649 IELFLKNAQLKDSIHVPNIYKQNKPEPYLYLVKKEVDKLEFIPKVKDMLKKEQAVLS 708
QY 694 TEGQSDNSEPSTEGETGOATTKPGQAGSALSGDSVQAQAEOKAQOPVPVPVPEAKA 753
DB 709 S-----ITQPLVAASETTEDGCHSTHTLSQSGETEVTETVGHGTTTITLPTQ- 758
QY 754 QVPTPPAPV-----NNKTENVSKLDYLEKLEIYELNTSYICHKYILYVSHSTPMNEK 802
DB 759 --PSPPEKVVVENSIEHKSNDNSQALTKTVYLLKLDELFTKSYICHKYILVSNSSMDQK 816
QY 803 ILKQYKITKEESKJSSCDPLDLFNIQNNIPVMYSFMFDSLNSLSQLPMEIYEKEMVCN 862
DB 817 LLEVNLTPPEENELKSCDPLDLFNIQNNIPAMYSLYDSMNDLQHLHFLFYKEMIY 876
QY 863 LYKLDNDKIKLLEAKKV-----STSVKTLSSSS-----MQPLSLTPQDK 904
DB 877 LHLKEENHIKKLEEQKQITGSTSTSSPGNTVTNTAQSAATHSNQSNQSNASSWTQNG 936
QY 905 PEVSAND---DTSHS---TNLNSLKLFPENILSLGSKNNIYQEL-IGOKSSENFYEKIL 956
DB 937 VAVSSGPAVVEESHDPVLTVLSISNDLKGIVSLNLGNKTKVNPPLTISTEMEKEFYENIL 996
QY 957 KDSDFYNESFTNFVKSADDINSLNDESKRKKLEEDINKLKTTLQLSFDLYNKKYKLE 1016
```



```
Db 1457 GELYKTHLAVNEEVKKVEADIKAEEDDKIKKIGSDSTKTEKTQSMKKAELKYLPPFLN 1516
QY 1420 NIETLYKTNDKIDLFVIHLEAKVLYTYEKSNNVEVKIKELNLYKTIOQKLADFKKNHF 1479
Db 1517 SLOKEYESLVSKVNTYTDNLKVINNCQLEKKEAEITVKKLDQYNKMDKLEBYKK---- 1572
QY 1480 VGIADLSTQYNNHNLTKFLSTGMPENLAKTVLSNLLDGNLQGMNLISOHOCVKKOCPO 1539
Db 1573 ---SEKNEVKSSGLEKMLKSLIKENESKEILSOLLNVQTLTMSSEHCIDTNVDP 1629
QY 1540 NSGCFRHLDERECKLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGNGKKI 1599
Db 1630 NAACYRYLDGMEWRCLLTFKEBGGKCVPGSNVTCCKDNGGCAPEACECKWTDSD---NKI 1685
QY 1600 TCCECTKPDSPYLPFDGIFCSCSNFLGIFSLILMLIL 1635
Db 1686 VCKCTKGESEPLFEGVFCSSSFLSLFLLMLLFL 1721

RESULT 9
Q9TYG1 PRELIMINARY; PRT; 539 AA.
AC Q9TYG1
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13358; BAA02619.1; -.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61075 MW; C70C2E100EC4A101 CRC64;

Query Match 33.3%; Score 2801; DB 5; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.8e-80;
Matches 539; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKKLSYLSGSLHHLIAELKEV 1160
Db 1 PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKKLSYLSGSLHHLIAELKEV 60

QY 1161 IKNKNTGNSPENNNDVNALESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV 1220
Db 61 IKNKNTGNSPENNNDVNALESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV 120

QY 1221 GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDGLGVVTVGEAVTPSIDNLTLSKIENE 1280
Db 121 GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDGLGVVTVGEAVTPSIDNLTLSKIENE 180

QY 1281 YEVLKPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKRENFKNVLESDLIPYKDLT 1340
Db 181 YEVLKPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKRENFKNVLESDLIPYKDLT 240

QY 1341 SSVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKLDLSI 1400
Db 241 SSVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKLDLSI 300

QY 1401 KKYINDKQGENEKYLPFLANNIETLYKTVDNKIDLFVIHLEAKVLYTYEKSNNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLANNIETLYKTVDNKIDLFVIHLEAKVLYTYEKSNNVEVKIKEL 360
```

```
QY 1401 KKYINDKQGENEKYLPFLANNIETLYKTVDNKIDLFVIHLEAKVLYTYEKSNNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLANNIETLYKTVDNKIDLFVIHLEAKVLYTYEKSNNVEVKIKEL 360
QY 1461 NYLKTIOQKLADFKKNNVFVGADISTDYNNHNLTKFLSTGMPENLAKTVLSNLLDGN 1520
Db 361 NYLKTIOQKLADFKKNNVFVGADISTDYNNHNLTKFLSTGMPENLAKTVLSNLLDGN 420
QY 1521 LOGMLNISQHCVKKOCPONSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGG 1580
Db 421 LOGMLNISQHCVKKOCPONSGCFRHLDERECKCLLNYKQEGDKCVENPNTCNENNGG 480
QY 1581 CDADAKCTBEDSGNGKKITCECTKPDSPYLPFDGIFCSCSNFLGIFSLILMLILYSFI 1639
Db 481 CDADAKCTBEDSGNGKKITCECTKPDSPYLPFDGIFCSCSNFLGIFSLILMLILYSFI 539

RESULT 10
Q25972 PRELIMINARY; PRT; 539 AA.
AC Q25972
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13360; BAA02621.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61076 MW; 72186F1412C57D1C CRC64;

Query Match 33.2%; Score 2798; DB 5; Length 539;
Best Local Similarity 99.8%; Pred. No. 2.2e-80;
Matches 538; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1101 PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKKLSYLSGSLHHLIAELKEV 1160
Db 1 PLKLTSEESIQTEDNYASLENFKVLSKLGKLDNLEKKKLSYLSGSLHHLIAELKEV 60

QY 1161 IKNKNTGNSPENNNDVNALESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV 1220
Db 61 IKNKNTGNSPENNNDVNALESYKKFLPEGTDVATVYSESGSDTLEOSQPKKPASTHV 120

QY 1221 GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDGLGVVTVGEAVTPSIDNLTLSKIENE 1280
Db 121 GAESNTITTSQNVDDDEVDDVVIIPFGESEEDYDGLGVVTVGEAVTPSIDNLTLSKIENE 180

QY 1281 YEVLKPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKRENFKNVLESDLIPYKDLT 1340
Db 181 YEVLKPLAGVYRSLLKQLENNVMTFNVNVDILNSRFNKRENFKNVLESDLIPYKDLT 240

QY 1341 SSVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKLDLSI 1400
Db 241 SSVVVKDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVLGYKILSEKYSKLDLSI 300

QY 1401 KKYINDKQGENEKYLPFLANNIETLYKTVDNKIDLFVIHLEAKVLYTYEKSNNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLANNIETLYKTVDNKIDLFVIHLEAKVLYTYEKSNNVEVKIKEL 360
```



```
|||||
Db 421 LQGLMLNLSHQCVKQCPQNSGCGFRHLDERECKLLNLYKQSGKCVENPNTCENNNG 480
QY 1581 CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLGIFLLILMLILYSFI 1639
Db 481 CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLGIFLLILMLILYSFI 539

RESULT 13
Q25976 PRELIMINARY; PRT; 539 AA.
AC Q25976;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13362; BAA02623.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
FT SEQUENCE 539 AA; 61045 MW; 8BF9C64322E9A778 CRC64;

Query Match 33.2%; Score 2794; DB 5; Length 539;
Best Local Similarity 99.8%; Pred. No. 3e-80;
Matches 538; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKTLSESIQTEDNVASLENFVKLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 1160
Db 1 PLKTLSESIQTEDNVASLENFVKLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKKNKNTGNSPSNNTDVNNALESYKFLPEGTDVATVVSSESDTLEQSQPKKPASTHV 1220
Db 61 IKKNKNTGNSPSNNTDVNNALESYKFLPEGTDVATVVSSESDTLEQSQPKKPASTHV 120
QY 1221 GAESNTITTSQNVDDVDDVIIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 1280
Db 121 GAESNTITTSQNVDDVDDVIIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 180
QY 1281 YEVLKPLAGVYRSKQLQLENNVMTFNVNVDILNSRFNKNFNKLVESDLIPYKDLT 1340
Db 181 YEVLKPLAGVYRSKQLQLENNVMTFNVNVDILNSRFNKNFNKLVESDLIPYKDLT 240
QY 1341 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSDLSI 1400
Db 241 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSDLSI 300
QY 1401 KKYINDKQGENEYLPFLNNIETLYKTVDNKIDLFIHLEAKVNLNLYTEKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEYLPFLNNIETLYKTVDNKIDLFIHLEAKVNLNLYTEKSNVEVKIKEL 360
QY 1461 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 1521 LQGLMLNLSHQCVKQCPQNSGCGFRHLDERECKLLNLYKQSGKCVENPNTCENNNG 1580
Db 421 LQGLMLNLSHQCVKQCPQNSGCGFRHLDERECKLLNLYKQSGKCVENPNTCENNNG 480
```

```
QY 1581 CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLGIFLLILMLILYSFI 1639
Db 481 CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLGIFLLILMLILYSFI 539

RESULT 14
Q25981 PRELIMINARY; PRT; 539 AA.
AC Q25981;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13363; BAA02624.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
FT SEQUENCE 539 AA; 61046 MW; 3ED87473EE87B65 CRC64;

Query Match 33.1%; Score 2791; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 3.7e-80;
Matches 537; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1101 PLKTLSESIQTEDNVASLENFVKLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 1160
Db 1 PLKTLSESIQTEDNVASLENFVKLSKLEGLKDNLNLEKKKLSYSSGLHLLIAELKEV 60
QY 1161 IKKNKNTGNSPSNNTDVNNALESYKFLPEGTDVATVVSSESDTLEQSQPKKPASTHV 1220
Db 61 IKKNKNTGNSPSNNTDVNNALESYKFLPEGTDVATVVSSESDTLEQSQPKKPASTHV 120
QY 1221 GAESNTITTSQNVDDVDDVIIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 1280
Db 121 GAESNTITTSQNVDDVDDVIIIPFGESEEDYDDLQGVVTVGEAVTPSVIDNLSKIENE 180
QY 1281 YEVLKPLAGVYRSKQLQLENNVMTFNVNVDILNSRFNKNFNKLVESDLIPYKDLT 1340
Db 181 YEVLKPLAGVYRSKQLQLENNVMTFNVNVDILNSRFNKNFNKLVESDLIPYKDLT 240
QY 1341 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSDLSI 1400
Db 241 SSVVVKDPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGYKILSEKYSKSDLSI 300
QY 1401 KKYINDKQGENEYLPFLNNIETLYKTVDNKIDLFIHLEAKVNLNLYTEKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEYLPFLNNIETLYKTVDNKIDLFIHLEAKVNLNLYTEKSNVEVKIKEL 360
QY 1461 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTIDQKLADFKKNNFVGIALSTDYNNHLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 1521 LQGLMLNLSHQCVKQCPQNSGCGFRHLDERECKLLNLYKQSGKCVENPNTCENNNG 1580
Db 421 LQGLMLNLSHQCVKQCPQNSGCGFRHLDERECKLLNLYKQSGKCVENPNTCENNNG 480
QY 1581 CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLGIFLLILMLILYSFI 1639
Db 481 CDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFIPLGIFLLILMLILYSFI 539
```


RESULT 15
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA;
AC Q25984;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE=93295445; PubMed=8515786;
RX Jongwutiwes S., Tanabe K., Kanbara H.;
RA "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RL field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13356; BAA02617.1; -;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61115 MW; 3788015F3127CB9E CRC64;
Query Match 33.1%; Score 2789; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 4.2e-80;
Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1101 PLKLTSEESIQTEDNVASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHHLIAELKEV 1160
Db 1 PLKLTSEESIQTEDNVASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHHLIAELKEV 60
QY 1161 INKNYTGNSPSNNTDVNNALESYKKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 1220
Db 61 INKNYTGNSPSNNTDVNNALESYKKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 120
QY 1221 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTVGEAVTPSIDNLSKIENE 1280
Db 121 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTVGEAVTPSIDNLSKIENE 180
QY 1281 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKFNKLVESDLIPYKDLT 1340
Db 181 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKFNKLVESDLIPYKDLT 240
QY 1341 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDTINFANDVLGYKYLSEKYSKSDLSI 1400
Db 241 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDTINFANDVLGYKYLSEKYSKSDLSI 300
QY 1401 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 360
QY 1461 NYLKTITQDKLADPKNNNFVGLADISTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTITQDKLADPKNNNFVGLADISTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 1521 LOGMLNISQHCVKKQCPONSGCFRHLDERECKCLLNNYKQEGDKCVENPNPTCNENNGG 1580
Db 421 LOGMLNISQHCVKKQCPONSGCFRHLDERECKCLLNNYKQEGDKCVENPNPTCNENNGG 480
QY 1581 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSSNFLGIFSFLILMLILYSFI 1639
Db 481 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSSNFLGIFSFLILMLILYSFI 539
RESULT 16

Q25971
ID Q25971 PRELIMINARY; PRT; 539 AA.
AC Q25971;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RX Jongwutiwes S., Tanabe K., Kanbara H.;
RA "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RL field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13359; BAA02620.1; -;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61017 MW; 5CA9C651BB62B5E1 CRC64;
Query Match 33.1%; Score 2788; DB 5; Length 539;
Best Local Similarity 99.6%; Pred. No. 4.6e-80;
Matches 537; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1101 PLKLTSEESIQTEDNVASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHHLIAELKEV 1160
Db 1 PLKLTSEESIQTEDNVASLENFKVLSKLEGKLDNLEKKKLSYLSGSLHHLIAELKEV 60
QY 1161 INKNYTGNSPSNNTDVNNALESYKKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 1220
Db 61 INKNYTGNSPSNNTDVNNALESYKKFLPEGTDTATVVSSESDTLEQSQPKKPASTHV 120
QY 1221 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTVGEAVTPSIDNLSKIENE 1280
Db 121 GAESNTITTSQNVDDVDVVIIVPIFGESEEDYDDLQGVVTVGEAVTPSIDNLSKIENE 180
QY 1281 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKFNKLVESDLIPYKDLT 1340
Db 181 YEVLYLKLPLAGYVRSLLKOLENNVMTFNVNVDILNSRFNKFNKLVESDLIPYKDLT 240
QY 1341 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDTINFANDVLGYKYLSEKYSKSDLSI 1400
Db 241 SSNYVVKDPYKFLNKEKRDKFLSSYNYIKDSITDTINFANDVLGYKYLSEKYSKSDLSI 300
QY 1401 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 1460
Db 301 KKYINDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLYNTEKSNVEVKIKEL 360
QY 1461 NYLKTITQDKLADPKNNNFVGLADISTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 1520
Db 361 NYLKTITQDKLADPKNNNFVGLADISTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDGN 420
QY 1521 LOGMLNISQHCVKKQCPONSGCFRHLDERECKCLLNNYKQEGDKCVENPNPTCNENNGG 1580
Db 421 LOGMLNISQHCVKKQCPONSGCFRHLDERECKCLLNNYKQEGDKCVENPNPTCNENNGG 480
QY 1581 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSSNFLGIFSFLILMLILYSFI 1639
Db 481 CDADAKCTEEDSGSNGKKITCCTKPDSPYPLFDGIFCSSSNFLGIFSFLILMLILYSFI 539
RESULT 17
Q25645
ID Q25645 PRELIMINARY; PRT; 1787 AA.
AC Q25645; 000883;

DT	01-NOV-1996 (TReMBLrel. 01, Created)
DT	01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE	Merozoite surface protein-1.
OS	Plasmodium berghei
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5821;
RN	[1]
RP	SEQUENCE OF 219-373 FROM N.A.
RC	STRAIN=K173;
RA	MEDLINE=97235961; PubMed=9080880;
RX	Toebe C.S., Clements J.D., Cardenas L., Jennings G.J., Wiser M.F.;
RT	"Evaluation of immunogenicity of an oral Salmonella vaccine expressing
RT	recombinant Plasmodium berghei merozoite surface protein-1.";
RL	Am. J. Trop. Med. Hyg. 56:192-199(1997).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K173;
RA	MEDLINE=98324404; PubMed=9662027;
RX	Jennings G.J., Toebe C.S., van Belkum A., Wiser M.F.;
RT	"The complete sequence of Plasmodium berghei merozoite surface
RT	protein-1 and its inter- and intra-species variability.";
RL	Mol. Biochem. Parasitol. 93:43-55(1998).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K173;
RA	Wiser M.F., Jennings G.J.;
RL	Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; U43521; AAC28871.1; -
SQ	SEQUENCE 1787 AA; 198156 MW; A585D64F51484EA CRC64;
	Query Match 30.1%; Score 2532; DB 5; Length 1787;
	Best Local Similarity 32.0%; Pred. No. 1.4e-71;
	Matches 605; Conservative 341; Mismatches 583; Indels 364; Gaps 41;
QY	1 MKIIFLGCSFLFIINTQCVTHESQVELVKKLEALEDAVLTCYSLFOKEKVLNEGTSGT 60
DB	1 MKVIGLLFSVFVFAICKKSEIEVYNDIIQKLESLSEGLELFQKSVQIIN----- 54
QY	61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSGVSGGSSRRRNP--SDNSSD 119
DB	55 -----ASPPSETINPFSDNT-- 69
QY	120 SDAKSYADLKHRVRYLLTIRKLYPOLFDLTNHLMTLCDNHGFKYLLIDGYEINELLY 179
DB	70 -----FAPKLGQFTIKFPELGFTQETELVNLIKLTLPVKPKGLKYLIESKEEFNGLMH 121
QY	180 KLNFFVDLLRAKLDVNCANDYCCQIPFNKIRANELDVLKLVFGYKRPDLNIDKNVCKME 239
DB	122 AINFYDVDFDKLDMCANNYCEIPEHLKISEETEMLKVKVILGYRPIENIQDIEKLE 181
QY	240 DYIKNKTTIENIELI--EESKVTIDNKK-----NATKEEKKKLYQAQYDLSTYNK 290
DB	182 IYITKNKETVTALNTLAEETKKTIPNEADCNDNTCDESKYSKKKIYQAMYNVIFYKK 241
QY	291 QLEEAHLNISVLEKRIIDLTKKNENIKELLOKLINEIKNPP-----PANSQ 334
DB	242 QLAELQVETLEKRVSAALKKNVDIKPPLQOIEDIAKAPVTTEGQITTSQSSTEPASTG 301
QY	335 NTPN----- 338
DB	302 -TPSSGEVSTGTSGGASAGVNTGAATTCTTGTAATCTTGAEAVTTCNTGAEAAATG 360
QY	339 -----TLLDNKKIEEKEIKEIAKTIKENIDSLFTDPLELYEYLREKN 383
DB	361 NTNTFVQVOTVPTLTPPEERKKKMDGLGYAIOIKETAKTIFKNLDGIFWNPIELEYLKEKK 420
QY	384 K---NIDISA--KVETKESTEP--NEYPNGVTVPLSVNDI-----NNALNELNSFGDLIN 431
DB	421 KESCNLTSSCKKNKTSETIIPLNVRYPNGIGVPLPNDVYNKIANNAAE--TTYGDLTN 478
QY	432 PFDTYKPSKNIVTDNRKFFINEIKIEKKIKIESDKKSYEDRSKSNLDTKGYEKL 491


```
QY 1428 VNDKIDLFVHLEAKVLYNTYKSNVEVKIKELNYLKTIOIDLADPKKN--NNFVGTADLS 1486
Db 1581 ILQSEEEYEVQLNRDLSYKNEKTEFEILTKMLEYIKIDEKLENFVENAENKHIASTA 1640
QY 1487 TDYNNHNLTKFLSGWPFENLAKTVLSNLLDGNLQGLMNLIS-OHQCVK-KOCPONSGCF 1544
Db 1641 L-----NNL-----NKSGLVGESEKILAKMLNMDSDLLGIDPKHVCINTRDIPANAGCF 1692
QY 1545 RHLDREBECKLLNTYKQBEDKCVENPNTPCNENNGCGCDADAKCTDEDSSGKKITCBCT 1604
Db 1693 RYDNGNEERWCLGKYKNNNTCIEDSNPTCGNNGGCDPTAGCQTAENRENSKKIITCK 1752
QY 1605 KDSYPLDFGICSSNPLGIGLFLILMLILYS 1637
Db 1753 EPTPNAYDGVFCSSSPFGLSILLIITLIVEN 1785

RESULT 18
Q9U6D4
ID Q9U6D4 PRELIMINARY; PRT; 1791 AA.
AC Q9U6D4;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1.
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=5823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ANKA;
RA MEDLINE=99330127; PubMed=10403333;
RX Zhong H., Fan J.Y., Yang S., Davidson E.A.;
RT "Cloning and characterization of the merozoite surface antigen 1 gene
of Plasmodium berghei.";
RL Am. J. Trop. Med. Hyg. 60:994-999 (1999).
DR EMBL; AF187232; AAF13063.1;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1791 AA; 198568 MW; 0474251901CC6A99 CRC64;

Query Match 30.0%; Score 2530; DB 5; Length 1791;
Best Local Similarity 31.8%; Pred. No. 1.7e-71;
Matches 604; Conservative 339; Mismatches 586; Indels 368; Gaps 40;

QY 1 MKIIFLCSFLFIINTQCVTHTESQELVKLEALEDAVLTCYSLFQEKVMNLECTST 60
Db 1 MKVIGLLFSVFFAICKSETIEVYNDIIIOKLEKLESUVEGLELFPKQSQVIIN----- 54

QY 61 AVTTSTPGSKGVSAGSGSGSVASGSGSVASGSGSGSGSRRTPN-SDNSSD 119
Db 55 -----ASPPSETINPESDNT--- 69

QY 120 SDAKSYADLUKRVNRYLLTIKELKYPQLPDLNHHMLTCDNTHGFKYLDGYEENELLY 179
Db 70 -----PAPLQGFITFEELGTEQTELVNLIKTLPVKNYGLKYLIESKEEFNGLMH 121

QY 180 KLNFEYDLRLAKNDVCANDYQIQIPENKIRANELDVLKLVGYRKPDLNIDKNVGMK 239
Db 122 ANFYDVRDLNDMCANNYCEIPEHLKISBEETEMLKVIILGVKRPENIQQDIEKLE 181

QY 240 DYIKNNKKTIENTINELI-BESKRTIDKNK-----NATKEEKKKLYQAQYDLSYKN 290
Db 182 IYITKNKEIVTALNTLIAETKKTIPENADCNDNDCTDESKYSKKKIIYQAMYNVIFYK 241

QY 291 QLEEAHNLVSLEKRIIDLTKKNENIKELLDKINEIKNPP-----PANGS 334
Db 242 QLAIEQKVIETLEKRVSAKKNVDIKPPLQIQIEDIKAAPVTTEGQITTSQSQSSTEPAG 301

QY 335 NTPN----- 338
Db 302 -TPSSGEVSTGTSGASAGVTNTGAATGTTGTGAATGTTTCAEAATGTTGCAEAATG 360
```

```
QY 339 -----TLLDNKKIEBEHEKEIAKTIKFNIDSLFTDPLELEYLYREKN 383
Db 361 NNTNTEVQVQVPTLTPEKKKKMDGLYQAQIKIETIKFNLDGIFVNFIELEYFKEKK 420
QY 384 K---NIDISA--KVETKESTEP--NEYPNGVTYPLSYNDI-----NNALNELNSFGDLIN 431
Db 421 KESCNLSTSSCKKNKTSETIIPLVRYPNGIGVPLPENDVYNKIANNAAE--TTYGDLTN 478
QY 432 PFDYTKEPSKNIYTDNERKKFNEIKEKIKIEKKKIESDKSYEDRSKSLNDITKYEKL 491
Db 479 P-DNTPITDELATNEQARKNLKAIKKKTEAEKNEKLESKTNYNDKLASFNQOKAPFKA 537
QY 492 LNEIYDSKNNNIDLTFEKMGMKRYKYVEKLTHTHTFASTYENSXHNLEKLTALKYME 551
Db 538 AKLFYESKFGNKLTSDFEKFKTQRTQTEYMNKYTELENCL--YGNTKQLSKLNKQLNYLQ 595
QY 552 DYSLRNIVVEKELKYKNIISKIENIETLVENIKKDEBQLEKFKITKDENRDEKILEV 611
Db 596 DYSLRKDIISDETEYFSNKKKGLQYNINRLAEAVQAKQNILVA-----SKDVPL 644
QY 612 SDIVKVOQKVLMMNKIDELKKTOLLKKNVELKHNHVPNSYKQENKQBPYYLIVLKKEI 671
Db 645 STLVELIQIKSLTLTKQIEQLNKTLSLRAQLKDKVYVPKSYGNGGKPEPYLIYAKREV 704
QY 672 DKLKVPMPKVESLINEEKNIKTE-----GOSDNSEPESTEGEITQATTKPGQQAGSAL 725
Db 705 DRLAQFIPAKIENNIKAKE--KTEQVPVVTGSEBETS-SVSTEVSVAQSSQSTSTVPA 761
QY 726 EGDS-----VOAQAEQEOKQAO-----PVPVPVPEAKAQVPTP-----P 759
Db 762 AGATSPVTVPVTEEAQSSQNAPTTAATPATTPPEAATTAATPATTPPEAATTTSTTTTS 821
QY 760 APVNNKTENVSKLDYLEKLYEELNTSYICHKYLIVSHSTMNEKILKQYKITKEES--KL 817
Db 822 TTTSTTTPVMTKLYLEKLOKFLVFSYCHKYLILLQNLSTINKDALSKYALTTEEDKIRTL 891
QY 818 SSCDPLDLLFNQNNIPVMYSMFDLSNLSQLFMIEYKEMVCNLYKLKD-NDKIKNL 876
Db 882 KRCSELDILLAIONNMPTWYSLYESIVDGLQNIYAELEYKEMHYIYKLDENPSIKSIL 941
QY 877 EEA-----KYSTSVKTSLSSSS 893
Db 942 VKAGVIDPEFVASPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1001
QY 894 MOPLSLTPODKPEVSANDDTSHSTLNLSLKLNFENILSLGKNKIYQELGOKSS----- 948
Db 1002 GASSETPASNPGASASSTPTPAAASNPGASASSTPTASNPASSTVQPTQVQPA 1061
QY 949 -----ENFYEKILKSDTFYNESFTNFVS 973
Db 1062 PAAPAAVPAQPADSNGSITTRAESIEDIPADDFELDNLYKSYLQRIDG-NNTEFINFIKS 1120
QY 974 KADDINLSNDESKRKKLEEDINKLTKTLQLSFDPLYKYLKLERLDFDKKTKVCKYKMQIK 1033
Db 1121 KKELIKALTPK-VKNOLYLEIAHLKELSHYDRIYKYLKLERLKEHEQIQVSRQRI 1179
QY 1034 KUTLLKEQLESKLSLNPNKHVLONFESVFENKKKEAIEAETENLTENTKILLKHYKGLVK 1093
Db 1180 ELSILKARLLKRONINGIFYILSGYVNFENKRRREADQYVDNALKNNDMLLYKARIK 1239
QY 1094 YINGESSPLKLTSEESIOTEDNYASLENFKVLSKLEGLKLDNLNLEKKKLSYLSGGLHL 1153
Db 1240 YFTSEAVPLKTLTKASLDRETNYLTKFRAYSRXELRFLKFNINLKESIYVSGGLYHV 1299
QY 1154 IAELEKVINKNYKNTGSPSENNTDNNALLESYKFLPEG-----TDVAT-- 1197
Db 1300 FEEFKELLKNKNTGKTNPDTPVEVNAFEQYKELLPGVITASASAAAATPTTSADAATQ 1359
QY 1198 -----VVYESGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDDVVIIPFGESE 1250
Db 1360 RATPESGSGSGSGSVSVSTPEEVAR--GSGENAVVSGSSVDD----- 1402
```



```
Qy 1251 EDYDGLGVVTEAVTPSVIDNHLKSIENYEVLYLKLPLAGVYRSLSKOLENNVMTFNV 1310
Db 1403 NDDDDIDQIASGQSENAQE-KNILEAPKNESEYLYARSIGITYLSLKHMTREFSTIKED 1461
Qy 1311 VKDILNSRFNFKNVWLESDLIPYKDLTSSNVVDDPYKFLNKRKRDRKFLSSYNYIKD 1370
Db 1462 MTGLNKKLQKRNDFLEVLNHELDFDLSTNKKYVIRNYPYLLDNDKKQIVNLKYYAK 1521
Qy 1371 SIDTD-----INFANDVLGYKILSEKYSKSLDSIKKYNKINDKOGENEKYLPLNMIET 1423
Db 1522 GVNEDIETAGIKFFNKKMIELYKIQAAVKEQIDAEEAATTKDK-EKKYVVPFEDLKG 1580
Qy 1424 LYKTVNDKIDFLVTHLEAKVNLTYEKSNEVEKIKELNYLTKTIQDKLADEKKN-NNEVGI 1482
Db 1581 LYETILGQSEYIEVLQNRDSYKNEKTEFEILTKNLETYIKIDEKLENEVENAENKHI 1640
Qy 1483 ADLSTYDNNHNLTKFLSTGWENFENIAKTVLSNLDGNLQGLNIS-OHOCVK-KOCPON 1540
Db 1641 ASIAL-----NNL-----NKSGLVGESESKILAKMLNDSMDLLGIDPKHVCINTROPAN 1692
Qy 1541 SGCFRHLDERECKCLLNKQEGDKCVENPNPFCNENNGCCDADAKCTEEDSGSNGKKIT 1600
Db 1693 AGCFRYDNGNEWRCLLGTGKNNNTCIEDSNPTCGNNGGCDPTAGCQTAENRENSKKII 1752
Qy 1601 CECTKPDSPYLFDFGIFCSSNFIIGISFLILMLILYS 1637
Db 1753 CTCKEPTNAYYDGVPCSSSFMGLSILLIILIVEN 1789

RESULT 19
Q25685 ID Q25685 PRELIMINARY; PRT: 1785 AA.
AC Q25685;
DC 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DS Major merozoite surface antigen precursor.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IPP-C1;
RX MEDLINE=91218805; PubMed=2090945;
RA Deleersnijder W., Hendrix D., Bendahman N., Hanegreets J., Brijs L.,
RA Hamers-Casterman C., Hamers R.;
RT "Molecular cloning and sequence analysis of the gene encoding the
RT major merozoite surface antigen of Plasmodium chabaudi chabaudi IP-
RT PC1.";
RL Mol. Biochem. Parasitol. 43:231-244(1990).
DR EMBL; M34947; AAA29730.1;
KW Merozoite; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1785 MAJOR MEROZOITE SURFACE ANTIGEN.
SQ SEQUENCE 1785 AA; 198887 MW; AC507F90113F4F7 CRC64;

Query Match 29.8%; Score 2509; DB:5; Length 1785;
Best Local Similarity 31.6%; Pred. No. 7.5e-71;
Matches 598; Conservative 324; Mismatches 605; Indels 366; Gaps 35;

Qy 1 MKIIFLCSPFFIINTQCWTHESYQELVKKLEALDEAVLTGYSLFQKEKWLNEGTSCT 60
Db 1 MKAIGLLFSVPFALYCKSETIGVNDLVHKLKLELSVEGLELFQKQSVIVNAQSPET 60
Qy 61 AVTTSTPGSKGVSAGSGGVSAGGVSAGGVSAGGVSAGGVSAGGVSAGGVSAGGVSAGGVS 120
Db 61 PVDPF-----TNP----- 68
Qy 121 DAKSYADLKHVRNRYLLTIKLYPOLFDLTNHLMTLCNINHGFYKYLIDGYEINELLYK 180
Db 69 -----EFAQKLQPFILKEELGFTQETELVNLIKTLGPNKYLGLYIESKEEFNLMHA 122
Qy 181 LNFYDILLRAKLNDCANDCYQIPFNKIRANDELVDLKKLVFGYRKPLDNIDNCKMED 240
```

```
Db 123 INFYDVLRLDKLDMCANNYCEIPEHLKINVEEIMLKKVILKVKRPENIODDLVLEEE 182
Qy 241 YIKKNNKTIENINELIBESKTKI-----DKNKNAKTEEEKKKLYQAQYDLSIYNKQ 291
Db 183 YIARKAKATAETLNLITEETKKITPEETDCNDTNCNDNTKYGKKKAIYQAMYNVIFYKKQ 242
Qy 292 LEEAHNLISVLEKRIDTLKKNNIKELLDKINEIKNPPANS----- 334
Db 243 LAELKKVIEVLEKRVATLKNEAIKPLQOQIEAIRGPPAVTEGQIATEGSSEETKQNSTE 302
Qy 335 --NTPNPLLD----- 342
Db 303 SSNTKTTTDDKAVTQATKATCTETNTGTETNTATGTTATGTTATGTTATGTTAT 362
Qy 343 -----NKKIEEHEKEIKEIAKTIKFNIDSLTFDPLEEYLYRE-NKNI 386
Db 363 EPVQVPAVQVLTBEERAKKIAELVYAKIKETAKTIKPNLDGIFVDPVELEYEYKKNESC 422
Qy 387 DISA---KVETKESTEP--NEYPNGVYPLS-----YNDINNALNELNSFGDLINPFDYTK 437
Db 423 HSTSSCHKANKTPTVPLNVRYPNGISYPLTEEVYSKIAHNAE--TTYGDLTN-VDNTA 480
Qy 438 EPSKNIYTDNERKFFINEIKEIKIEKKIESDKKSYEDRSKSLNDITREYEKLLNEIYD 497
Db 481 ITEDLTNEQARKNLIIKAIKKIEAEQKLVKDDYDTKLAFAFNQGTPTFKAAKKFFE 540
Qy 498 SKFNNIDLTNFEKMMGKRYSYKVEKLTHTHTFASYSKSHNLEKTKALKKMYEDSLRN 557
Db 541 SKFRNKLTDDIEDDFKTKRTEYMNKAA--LVGCEYGTQQLINKLNKQLNYLDGLRK 598
Qy 558 IVVEKELKYKNIKLSKIENEIETLVENIKKDEQLEFKKITDKENKPKDEKILVSDIVK 617
Db 599 EIVNTEIEYFSNKKSELOYNINRLANAVQAQONILVA-----SKHPLSLTVEL 647
Qy 618 QVQVLLMKNKIDELKKTQLILKNVELKUNIHVPNSVKQENKQEPYVYLVLEKIDKLVK 677
Db 648 QIQKSLTLKLIBQLNKTFFESLNKHLKDKIYVPQYIGKEGPEPYLIAIKKIEDRLAKF 707
Qy 678 MPKVESLINEEKKNIKTEG-QSDNSEPSTEGEITQOATTKPGQQA----- 721
Db 708 IPKIDDMIEKEKQMEQEHVATGESEQASSAGTSSTETTSOTAPAVPAAPAPAEKAKE 767
Qy 722 -----GSALEGDSV---QAQAEQOKQAP-----PVPVPEAKAQVPT 757
Db 768 GTESTEETPAASKPAEGAAAGTATPTTEQEAAPTEQEAQAPAEVPAETPAAPT 827
Qy 758 PPAPVNNKTENVSKIDYLEKLYEFLNTSYICHKYILVSHSTMNEKILKQYKITKEES-- 815
Db 828 PAAPAPAKPVMTKLYVLEKLLKFLAFSYACHKYVLLQNSTINKDALSXYALTPEEDKIR 887
Qy 816 KLSSCDPLDLLFNQNNIPVWYSMFDLSNLSOLFMEIYEKEMVCNLYKLKD-NDKTKN 874
Db 888 TLKRCSELDVLLATQNNMPTMYSLYENVVGLQNIYTELYEKENMYHIYNLKNKPAVK 947
Qy 875 LL-----EEAKKVSTSVKT- 888
Db 948 LLVAGVIDPEVPAPTPAVPAPETAPETAPETAPETAPETAPETAPETAPETAPET 1007
Qy 889 LSSSSMPLSLTPQDKP-----EVSANDDTSH 915
Db 1008 AESASTEPTKAPTATPTSETVTEGTTTAPAKQAEQASSAPAPAKPAQVTVTGQ 1067
Qy 916 STNLNLSKLFEINILSLGKNKNNTYQELIGOKSSNFYEKILKDSDTFYNESTFNVSKA 975
Db 1068 STNVEGSTQV-----RAESEDMEFVDDEVDNFYKSVLQQVDG-NNTQFIDFISKK 1118
Qy 976 DQINSLNDSKRRKLBEDINKLKTQLQSLFDLYNKKYKLERLFDKKTKTVGKYKMOIKKL 1035
Db 1119 ELINALTPE-KVNQLYDLTAHLKEUSEHYNRYNRYKYLKLERLYQHQEQTEAANKVKEI 1177
Qy 1036 TLLKEQLESKLSLNNPKRHVLQNFVSFFNKKKEAEIETENTLENTKILLKHVGLVKY 1095
Db 1036 TLLKEQLESKLSLNNPKRHVLQNFVSFFNKKKEAEIETENTLENTKILLKHVGLVKY 1095
```


Db 1178 SVLSRLKRRKKYINGTFYVLSGFANFNKRREAEKQYVDNAIKNTDMLLKYKARSYF 1237
Qy 1096 NGESSPLKTLSEESIQOTEDNVASLENKSVLSKLGKLDNLNLNKKKLSLYSSGLHLIA 1155
Db 1238 TSEAVPLKTLTKSIDREANVLTKEFRAYSRLRLKKNLNLKGRITVSGGLHVF 1297
Qy 1156 ELKEVIKNKNTYGNPSNNNTVNNALESYKFLPEGTD-----VATVSESGSDTLEQS 1210
Db 1298 EFKELLNKGTGTNPENAPEVKAPEQYKELLPGATTAPVVPVAPAPATAAPAA 1357
Qy 1211 QPKP-----ASTHVGAESNTITTSQNVDDVDDVVIIPFGSEEDYDGLQ 1258
Db 1358 DAPVAAAAAASGSAATTEGAATTVVASSDND-----DDDDMDQ 1402
Qy 1259 VVTGEAVTPSIDNLSKIEYEVLYLKLPLAGVYLSKLENNVMTFNVNVKDIINSR 1318
Db 1403 IANAQS--TDEVKDILDFAKSENEYIYKSLGNTYKSFKKHMLKEFSMIKEDINTGLNYK 1461
Qy 1319 FNKREPNKLVESLIDIPKDLTSSNVVYKDYKFLNKEKRDKFLSSYNYKDSIDTDINF 1378
Db 1462 LEKRNDFDLVSYELAFKDLINTNKFVKNPYQLLDNDKDKQMINLKYAIKGVTEDET 1521
Qy 1379 ANDVLGYKILSEKYSKSDLSIKYI-----NDKQGENEKYLPFLNLIETLYKTVDNKID 1433
Db 1522 ATDGIIEFNKMIELYKPOLNAVNQIAAIGTEPTDAEKKKIYAPIFEDLKLGLYETILNGAE 1581
Qy 1434 LFVTHLEAKVLYNYTEKYSNVEVLIKELNYLKTIODKLADF----KNNNFVGIADLSTDY 1489
Db 1582 EFSELLOHKLNYKIERAGEDILMANLEYIRIDEKLEDFVESAEKKNHTASIA-----1635
Qy 1490 NHNLLKFKLSTGMVFENLAKTVLSNLDGNLQGLMNI--SOHQCVKQKOPNSGCFRHL 1548
Db 1636 -----LNNLNSGLVTEGESKKILAKMLNDAMDLLGTSGNHVHCISTSTPDNAGCFRYDD 1690
Qy 1549 ERECKCLLNYKQ--EGDKCVENPNPCNENNGCGDADAKTEEDSGSN--GKKITCECT 1604
Db 1691 GTEWRLLGFKDDGDNRCVADAPVCNNNGCGDNKADCREVENTDRDPSKKIVCTCK 1750
Qy 1605 KPDSPYLFDDGIFCSSNFGISFLILMLILYS 1637
Db 1751 EPNNAVYAGVFCSSGFMGLSILLITLIVEN 1783

RESULT 20

Q25668 PRELIMINARY; PRT; 1766 AA.
AC Q25668;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein-1 precursor.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS;
RX MEDLINE=94187794; PubMed=7511214;
RA McKean P.G., O'Dea K., Brown K.N.;
RT "Nucleotide sequence analysis and epitope mapping of the merozoite
surface protein 1 from Plasmodium chabaudi chabaudi AS.";
RL Mol. Biochem. Parasitol. 62:199-210(1993).
DR EMBL; L22982; AAA29499.1; -;
KW Signal.
FT SIGNAL.
FT CHAIN 1 19 POTENTIAL.
SQ SEQUENCE 1766 AA; 197557 MW; 6CFE98E26EDBD4C CRC64;

Query Match 29.6%; Score 2494.5; DB 5; Length 1766;
Best Local Similarity 32.0%; Pred. No. 2.1e-70;
Matches 603; Conservative 334; Mismatches 581; Indels 365; Gaps 43;

Qy 1 MKIIFLCFLFFINTQCVTHTESYQSLVKKLEALEDAVLGTYSLFOKERMWLNESGTS 60

Db 1 MKAIGLFSFVFFAIYCKSETIGVYNDLVHKLLEELSELSVEGLFOKSOVIVNAQSPT 60
Qy 61 AVTTSTPGSKSVASGGSGSVASGGSVASGGSVASGGSGNSRRTPNPSNDS 120
Db 61 -----PVPDFTDS 68
Qy 121 DAKSYADLKHVRNRYLLTIKELYPQLFDLTNHHMLTLCODNIHGFKYLDGYEENELLYK 180
Db 69 -----NFAPKLOEFTDFEELGTEQKELVNLKLGPNKYGLKYLESKEEFNLMHA 122
Qy 181 LNFYDILLRAKLDNVANDYCOIPFNKIRANELDVLKLVFGYRKPLDNLKDNVGMKED 240
Db 123 INFYDVLDRKLDNMCANNYCEIPEHLKINVEIEMLKVVGLYRKPTENIQDVLKLEE 182
Qy 241 YIKKNTKIENINELI-----ESKKTIDKNKNATKEEKKKLYQAOVDLSIYNKQ 291
Db 183 YIANKKATATLNNLIETEAKITPEEAECDNTCTNDKYKKKAIQAMVNYFYKKQ 242
Qy 292 LEEAHLISVLEKRDITLKKNNIKELLDKINEKNP-----328
Db 243 LAEIKKVLEVLEKRVATLKKNEAIKPLLLQLEAIRGPATVTEGQITTEGTEETKQNDAA 302
Qy 329 -----PPANSNT-----PNTLL-----DNKKI 347
Db 303 QTATKTTTTGSGAGANTNETANAGATATVATGTATNTVTVQVPPVPLTEEKAKKI 362
Qy 348 EEEKEKEIKIAKTIKFNIDSLETPLEYYLRE--KKNIDISA---KVETKESPEP--N 401
Db 363 AELYAQKEIAKTIKFNLDGIFVDPVELEYKKEKKNESCHSTSSCHANKKPTETVPLNV 422
Qy 402 EYPNGVTYPLS-----YNDINNALNELNSFGDLINPFDYTKEPSKNIYTDNE--RKKFINE 455
Db 423 RYPNGISYPLTEEVYYSKIAHNAAE--TTYGDLTN--VDNT--PITEDLTNEQARKNLKA 478
Qy 456 IKKIKTEKKIESDKSYEDRSKLDNITKEYEKLNEIYDSKFNNDIDLTNPEKMMGK 515
Db 479 IKKIEAEQKLVTKADYDTKLAEFNGQKTPFKEAAKNFYESFRNKLTTETFEAFKTK 538
Qy 516 RYSYKVEKLTHNTFASYENSKHNEKLTAKLYMEDYSLRNIVVEKELKYKNLISKIE 575
Db 539 RTEYMNKAA--LVGCEYGNTOOLINKLNKQNLVLDYGLRKEIVNTEIEYFSNKKSELO 596
Qy 576 NEIETLVENIKKDEQLPEKKITKDNKDPDEKILEVSDIVKQVQVLLMMKIDELKTKQ 635
Db 597 YNINRLANAVQAKQNLVA-----SKHIPSLTLVQLQIKSLTLKLEQNLNTE 645
Qy 636 LILANVELKHNHVPNSYKQENKOEPPYVLYLVKKKEIDKLKVFMPKVESLINEEKNIKTE 695
Db 646 FSLNKAHLKDKIYVPQTYGKRGKPEPPYVLYIAIKKEIDRLAKFIPKIDDMIKEKQKMEQ 705
Qy 696 ----GQSD-----NSEPSTEGEITGQATTGQOAGSALSGDSVQAQAOEQ-----737
Db 706 PVATGESEQVTTSSGTVTSTQSAQAPTPPAAPAPAKAEGAETTEKAKEDITETKEPTAT 765
Qy 738 -----KQAP--PVPVVPPEAKAQVTPPPAPVNNKNTENYSKLDYLEKL 778
Db 766 EETSETPAAAPETTPVTEEAQAPAAEVPAPTTPAAPAPAAAPAK---PVMTKLYLEKL 822
Qy 779 YEPINTSYICHKYLVSHSTWNEKILQYKITEEES--KLSSCDPLDLLFNQNNIPVM 836
Db 823 KKFLAFSYACHKYVLLQNSTINKDALSKYALTPEEDKIRTILKRCSELDVLLAIGNMPTM 882
Qy 837 YSNFDSLNSLSQLFMEIYEKEMCNLYKLDK--NDKIKNLL-----876
Db 883 YSLYENVVDGLQNIYTELEKEMMYHIYNLKDKNPAVKALLVKGAVTEPEPEPVPPTPAV 942
Qy 877 -----EEAKKVTSTVKTLSSSMQPSLSTPQDKPEVSAND-----912
Db 943 PAPETAPVQEAQQTQETTQETTPGTTPETATPGTTPETTTTPEAPAEPAENASVTVQEGTTP 1002
Qy 913 -----TSHSTLNNSLKLLENIL 930

Db 1003 AEAPAAQKQAGASAPSAETVPAPAPAPQVTSQASTQVSGQSTNGEGNTKV----- 1057
QY 931 SLGKNKIYQELIGQKSSSENFYEKILKSDTFYNESFTNFVSKABDINSLEDSKRKKL 990
Db 1058 ---SAESEDEMFDDVEVDNFYSYLLQVDG--NTQFIDIFIKSKKLINALTPK-KVNQL 1112
QY 991 BEDINKLKTQLSFDLYKYLKRLRFDKKTGVGYKQMKIKLTLLEQLESKLSLN 1050
Db 1113 YLDIAHLKLESHYNYRYKYKLRLYOKHEQIEAANKVKEISVLKSRLLKRRKYIN 1172
QY 1051 NPKHVLQNFSEFFNKKEAEIAETENTLTKLLKHXYGLVYKNGESSPLKTSEESI 1110
Db 1173 GTFYVLSGFANFNKREAEQYVNAIKNTDMLKYKARSKYFTSEAVPLKTLTKTSI 1232
QY 1111 QTEDNYASLNFVSKLGLKGLKDLNLKELKLSYLSGLLHLLAELEKVIKNKNTGNS 1170
Db 1233 DREANYLKEKFRAYSRLRLKLNINLGKERITVSGGLHVFEEFKLLANKNGTGT 1292
QY 1171 PSENNTVNNALSKKFLPEGTD-----VAPVSESGSDTLEQSQP-----KKPAS- 1217
Db 1293 NPENAPEVIKAEQYKELLPGATTAPVPTPAVAPASATTPADEPAAAAAASAPASG 1352
QY 1218 --THVGAESNT--ITTSQNVDDVDVVIIPFGESEEDYDLGQVVTGEAVTPSVIDNI 1273
Db 1353 SATTTGGGCTPVANSNDOD-----DDDDMDQIANAQSS--TDGEVKDI 1396
QY 1274 LSKINEVEVLYKPLAGVYLSLKLQLENNVMTFNVNVDI---LNSRFNKRNFKNVLE 1330
Db 1397 LDEFKSENEYTVKSLGNTYKSEKHM---LKEFSMIKEDIITGLNYKLEKRNDFLDVLS 1453
QY 1331 SDLIPYKDLTSSNVVYKPYKFLNKEKDKFLSSVNYTKDSIDTIDINPANDVLGYKILS 1390
Db 1454 YELALFKDINTKFKVNPYQLLNDKDKQMINLKYAIGVTEDETATDGCIEFFNKW 1513
QY 1391 EYKYSOLD-----SIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVHLEAKV 1443
Db 1514 ELYKPOLNAVNEQIAIEKETDKE-EKKKYVPFEDLGLUYETILNGAEFESELLQHL 1572
QY 1444 LNYTEKSNVEVKELNYLKTIOKLADE---KKNNFVGIADLSTDYNNHNLTKFL 1499
Db 1573 ENYKIEKAGFDILMANLEYIRIDEKLEDFVESAEKNKHASIA-----LNNLN 1621
QY 1500 STGMVFENLAKTVLSNLLDGNLQGLMNT-SHQCVKQKQCPQNSGCFPHLDERBECKCLLN 1558
Db 1622 KSLGVTGESKKILAKILNDMDALLGIGSNHVCIGTNIPENAGCFRYDGGKEWRCLLG 1681
QY 1559 YKQ--EGDKCVENPNPTCNENNGGCDADAKTEEDSG--SNGKKITCECTKPDSPYPLFDG 1614
Db 1682 FKNNDGTRCEKDEATCSINKGCDPSAECREVENADRENSKKNVCTCPTNAYIDG 1741
QY 1615 IFCSSSNFLGISFLILLMLILYS 1637
Db 1742 VFCSSSGFMGLSILLITLIVFN 1764

RESULT 21

Q25924
ID Q25924 PRELIMINARY; PRT; 651 AA.
AC Q25924;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71;
RX MEDLINE=92275047; PubMed=1592091;
RA Olafsson P., Matile H., Certa U.;
RT "Plasmodium falciparum: the repetitive MSA-1 surface protein of the
RO-71 isolate is recognized by mouse antibody against the

RT nonrepetitive repeat block of RO-33.";
RL Exp. Parasitol. 74:381-389(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RO-71;
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; Z35329; CAA8458.1; -;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
SQ SEQUENCE 651 AA; 74134 MW; AA2137E699255150 CRC64;

Query Match

Best Local Similarity 25.5%; Score 2146; DB 5; Length 651;
Matches 422; Conservative 82; Mismatches 114; Indels 36; Gaps 5;

QY 1019 FDKKTVGYKQMKIKLTLLEQLESKLSLNPNKHVLFQNFVFNKKKEAEIAETENTL 1078
Db 1 FNKKKELGQDKMQIKKLTLLKEQLESKLSLNPNVNFQNFVFNKKKEAEIAETENTL 60
QY 1079 ENTKILLKHXYGLVYKNGESSPLKTSEESIQTEDNTASLENFVLSKLEGLKDLNML 1138
Db 61 ENTKILLKHXYGLVYKNGESSPLKTSEESIQTEDNTANLEKFRALSKIDKLDNHL 120
QY 1139 EKKLSYLSGLLHLLIAELKEVKNKNTGNSPSNNTDNNALSYKFLPEGTDVATV 1198
Db 121 GKKLSPSSGLHLITELKEVINKNKNTGNSPENNNKVKNEALUSYENFUE- AKVTV 179
QY 1199 VSE-----SGSDTLEQSQPKKPASTHVGAESNTITTSQNVDDVDVIVP 1244
Db 180 VTPQPDVTPSPLSVRVSGSGSTKEETQIPTSGSLTLEQVQVQLQNYDEEDSLVLP 239
QY 1245 IFGESEEDYDLGQVVTGEAVTPSVIDNLSKIENEVEVLYKPLAGVYRSLLKQLENNV 1304
Db 240 IFGESEENDEYDQVVTGEAISVT-MDNILSGFENEYDVYIYKPLAGVYRSLLKQIEKI 298
QY 1305 MTFNVNVDILNSRFNKRNFKNVLESILPYKDLTSSNVVVKDYKFLNKEKDKFLSS 1364
Db 299 FTFNLNLDILNSRLKRYFLDVLSDLMQFKHISSEYIIESFKLLNSEQNTLKS 358
QY 1365 YNVIKDSIDTIDINPANDVLGYKILSEKYSKSDLSIKKYI-----N 1405
Db 359 YKIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKKVIKEKEEFPSPPTTSPAKT 418
QY 1406 DKOGENEKYLPFLNNIETLYKTVNDKIDLFVHLEAKVLYTEKSNVEYKIKELNYLKT 1465
Db 419 DEQKESKFLPFLTNITETLYNNLVNKIDDLINLAKINDCNVEKDEAHVKITLSDLKA 478
QY 1466 IQDLADFKKNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQGLM 1525
Db 479 IDDKIDLFKNHNDFEAIKLLINDTKKMLGKLLSTGLV-QNFPNTIISKLEGGKFDML 537
QY 1526 NISQHCYKQCPQNSGCFPHLDERBECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 1585
Db 538 NISQHCYKQCPQNSGCFPHLDERBECKCLLNKQEGDKCVENPNPTCNENNGGCDADA 597
QY 1586 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSSNPLGLISFLILLMLILYSFI 1639
Db 598 KCTEEDSGSGNGKITCECTKPDSPYPLFDGIFCSSNPLGLISFLILLMLILYSFI 651

RESULT 22

Q25923
ID Q25923 PRELIMINARY; PRT; 652 AA.
AC Q25923;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface antigen 1 (Fragment).

OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=88142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of
RT Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=FCH5/NF7;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1";
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; 235328; CAA84557.1; -;
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
FT NON_TER 1
SQ SEQUENCE 552 AA; 74292 MW; 2B6A87737B490A62 CRC64;

Query Match 25.2%; Score 2119.5; DB 5; Length 652;
Best Local Similarity 63.8%; Pred. No. 3.8e-59;
Matches 418; Conservative 84; Mismatches 116; Indels 37; Gaps 5;

QY 1019 FDKKTVGKYKMQIKKLTILKEQLSKLSLNNPKHVLQNFSPVFNKKKAEIAETNTL 1078
DB 1 FNKKELGQDKMQIKKLTILKEQLSKLSLNNPHNVLQNFSPVFNKKKAEIAETNTL 60

QY 1079 ENTKILLKHVGLVYKNGSSPLKLTSESIOTEDNYASLENFKVLSKEGKLKDNLL 1138
DB 61 ENTKILLKHVGLVYKNGSSPLKLTSESIOTEDNYANLEKFRVLSKIDGKLNLDLHL 120

QY 1139 EKKKLSYSSGLHHLIAELKEVKNKNTGNSPSENNTDYNNALESYKFLPEGTVDATY 1198
DB 121 GKKLSPSSGLHHLITELKEVKNKNTGNSPSENKKVNEALKSVENFLPE-AKVTY 179

QY 1199 VSE-----SGSDTLEQSQPKPKPASTHVGAESENTTTSQNVDEDDVVIIP 1244
DB 180 VTPPQDPVTPSLSVRSVSGSGSTKEETQIPTSGSLLELQVQVQSQNYDEEDSLVLP 239

QY 1245 IFGESEDDYDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKLPGVYRSLLKOLENNV 1304
DB 240 IFGESEDDYDLGQVVTGEAVT-NDNLSGFENEYEVLYLKLPGVYRSLLKOLENNV 298

QY 1305 MTFNVNVDILNSRFNKNRKNFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKRDRFLSS 1364
DB 299 FTENLNLNDILNSRLKRRYFLVLESDLQMFQKHSSNEVYIIEDSPKLLANSEQKNTLLKS 358

QY 1365 YNTIKSIDTDINFANDVGLYKILSEKYSKSDLSIKKYI----- 1404
DB 359 YKIKESVENDIRFAQEGISYEVKYLAKYKDDLESIKKVIKEKEKFPSPPTTPSPAK 418

QY 1405 NDKOGENEKYLPLFNNLETLYKVNKDILFVHLEAKVLNYYEKSNNVVKIKELNYLK 1464
DB 419 TDSQKESKFLPPLTNETLYNNLVNKKIDDYLLNLAKINDCNVCEKDEAHVKTIKLSDLK 478

QY 1465 TIQDKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQGM 1524
DB 479 AIDKIDLFKNPYDFAIKKLINDTKKMLGKLLSTGLV-QNFPNTIISKLEIGKFDQM 537

QY 1525 LNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDAD 1584.
|||||

DB 538 LNTSQHOCVKKQCPQNSGCGFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDAD 597
QY 1585 AKCTEEDSGSGNGKKTCECTKPDSPYPLFDGIFCSSNPLGTSFLLILMLILYSFI 1639
DB 598 ATCTEKDGSGRKKTCECTKPDSPYPLFDGIFCSSNPLGTSFLLILMLILYSFI 652

RESULT 23
Q03999 PRELIMINARY; PRT; 400 AA.
AC Q03999;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 precursor (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RT "Proteolytic processing of the Plasmodium falciparum merozoite surface
RT protein-1 produces a membrane-bound fragment containing two epidermal
RT growth factor-like domains";
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; M64681; AAA29709.1; -;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Signal.
FT NON_TER 1
FT SIGNAL <1 24 POTENTIAL
FT CHAIN 25 286 MEROZOITE SURFACE PROTEIN 1.
FT CHAIN 287 400 MEROZOITE SURFACE PROTEIN 1.
SQ SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;

Query Match 24.9%; Score 2101; DB 5; Length 400;
Best Local Similarity 100.0%; Pred. No. 8.8e-59;
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 VIIVPIFGESEDDYDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKLPGVYRSLLKQ 1299
DB 1 VIIVPIFGESEDDYDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKLPGVYRSLLKQ 60.

QY 1300 LENNVMTFNVNVDILNSRFNKNRKNFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKEKD 1359
DB 61 LENNVMTFNVNVDILNSRFNKNRKNFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKEKD 120

QY 1360 KFLSSYNYIKSDITDINFANDVGLYKILSEKYSKSDLSIKKYINDKQGENEYLPPLN 1419
DB 121 KFLSSYNYIKSDITDINFANDVGLYKILSEKYSKSDLSIKKYINDKQGENEYLPPLN 180

QY 1420 NIETLYKTVNDKIDLFIHLEAKVLNYYEKSNNVVKIKELNYLKTIOKLADEKKNNF 1479
DB 181 NIETLYKTVNDKIDLFIHLEAKVLNYYEKSNNVVKIKELNYLKTIOKLADEKKNNF 240

QY 1480 VGTADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQMLNLSHQCVYKQCPQ 1539
DB 241 VGTADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGNLQMLNLSHQCVYKQCPQ 300

QY 1540 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADACTEDSGSNCKI 1599
DB 301 NSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADACTEDSGSNCKI 360

QY 1600 TCCTKPDSPYPLFDGIFCSSNPLGTSFLLILMLILYSFI 1639
DB 361 TCCTKPDSPYPLFDGIFCSSNPLGTSFLLILMLILYSFI 400

RESULT 24
Q9BMG8


```

ID Q9BMG8 PRELIMINARY; PRT; 376 AA.
AC Q9BMG8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCB-1;
RA Li X.R., Goel V.K., Liu S.C., Chishti A.H., Oh S.S.;
RT "42 kDa subfragment of MSP-1 gene of Plasmodium falciparum FCB-1 isolate."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF325919; AAK07641.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 376 AA; 43218 MW; EB2A72EDF231A5AF CRC64;

Query Match 23.5%; Score 1979; DB 5; Length 376;
Best Local Similarity 100.0%; Pred. No. 5.2e-55;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1264 AVTPSVINILSKTENEVEVLYLPLAGVYRSKQLQENNVMTFNVVKDILNSRFNKR 1323
Db 1 AVTPSVINILSKTENEVEVLYLPLAGVYRSKQLQENNVMTFNVVKDILNSRFNKR 60

Qy 1324 NFKNVLESDLPYKDLTSSNVVVDPKYKFLNKRDRKFLSSYNYIKSDIDTDFANDVL 1383
Db 1 NFKNVLESDLPYKDLTSSNVVVDPKYKFLNKRDRKFLSSYNYIKSDIDTDFANDVL 120

Qy 1384 GYYKILSEKYSDDLSTKYYINDQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKV 1443
Db 121 GYYKILSEKYSDDLSTKYYINDQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKV 180

Qy 1444 LNYTYEKSNEVKIKELNYLKTIDQKLADFKKNNFVGADLSTDTYNNHLLTKFLSGM 1503
Db 181 LNYTYEKSNEVKIKELNYLKTIDQKLADFKKNNFVGADLSTDTYNNHLLTKFLSGM 240

Qy 1504 VFENLAKTVLSNLLDGNLQGLMNIHQCVKKQCPQNSGCFRHLDEREECKLLNYKQEG 1563
Db 241 VFENLAKTVLSNLLDGNLQGLMNIHQCVKKQCPQNSGCFRHLDEREECKLLNYKQEG 300

Qy 1564 DKCVENPNTCNENGGCDADAKTEEDSGSGKGIKTCCTKPDSPYPLFDGIFCSSNFL 1623
Db 301 DKCVENPNTCNENGGCDADAKTEEDSGSGKGIKTCCTKPDSPYPLFDGIFCSSNFL 360

Qy 1624 GISFLLILMLILYSFI 1639
Db 361 GISFLLILMLILYSFI 376

RESULT 25
Q9NAT3
ID Q9NAT3 PRELIMINARY; PRT; 360 AA.
AC Q9NAT3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC Goel V.K., Liu S., Chishti A.H., Oh S.S.;

RT "38 kDa subfragment of MSP-1 gene of FCB-1 strain of Plasmodium falciparum."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF286876; AAF67595.1;
DT NON_TER 1
FT NON_TER 360
SQ SEQUENCE 360 AA; 40993 MW; 9A5E9DE65E5A680B CRC64;

Query Match 21.6%; Score 1818; DB 5; Length 360;
Best Local Similarity 99.7%; Pred. No. 5.2e-50;
Matches 359; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 902 QDKPEVSANDDTSHSTNLNLSKLFENILSLGKKNKIYOELIGOKSSSENFYEKILKSDT 961
Db 1 QDKPEVSANDDTSHSTNLNLSKLFENILSLGKKNKIYOELIGOKSSSENFYEKILKSDT 60

Qy 962 FYNESFTNFVSKADDDINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLERLFDK 1021
Db 61 FYNESFTNFVSKADDDINSLNDESKRKKLEEDINKLKTQLSFDLYNKYKLERLFDK 120

Qy 1022 KKTGVKYMOKIKLTLLEQLESKLSLNPNKHVLQNFVFNNKKEAEIAETENTLENT 1081
Db 121 KKTGVKYMOKIKLTLLEQLESKLSLNPNKHVLQNFVFNNKKEAEIAETENTLENT 180

Qy 1082 KILKHKYGLVYKNGESSPLKLTSESIQTDENYASLENFKVLSKLGKLDNUNLEKK 1141
Db 181 KILKHKYGLVYKNGESSPLKLTSESIQTDENYASLENFKVLSKLGKLDNUNLEKK 240

Qy 1142 KLSYSSGLHLLIAELAEVIRKNKNTGNSPENNTDNNALSKYKFLPEGTDVATVSE 1201
Db 241 KLSYSSGLHLLIAELAEVIRKNKNTGNSPENNTDNNALSKYKFLPEGTDVATVSE 300

Qy 1202 SGSDTLEQSQPKPASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDDLQGVVT 1261
Db 301 SGSDTLEQSQPKPASTHVGAESNTITTSQNVDDVDVVIIPFGESEEDYDDLQGVVT 360

RESULT 26
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1;
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 21.1%; Score 1778; DB 5; Length 569;
Best Local Similarity 60.8%; Pred. No. 1.5e-48;
Matches 348; Conservative 79; Mismatches 109; Indels 36; Gaps 5;

Qy 1101 PLKTLSESIQTDENYASLENFKVLSKLGKLDNUNLEKKLSYSSGLHLLIAELKEV 1160
Db 1 PLKTLSESIQTDENYANLEKFRVLSKIDGKLDNHLGKKKLSYSSGLHLLIAELKEV 60

Qy 1161 IKKNYTGNSPENNTDNNALSKYKFLPEGTDVATVSE-----SGSDT 1206
```


Db	61	IKKNYTGSPSENKKVNEALKS	YENFLPE-AKVT	VTVPQDPVTSPLSVRVSSGG	119			
Qy	1207	LEQSQPKKPASTHVGAESNT	ITTSQNVDDVDVVIIP	IFGSEEDYDGLQGVVTG	1266			
Db	120	STKEETQIPTSGSLLTELQ	QVVLQYDEEDSLVPL	IFGSEEDNDEYLDQVVT	179			
Qy	1267	PSVIDNTLSKIENEYEV	ILKPLAGVYRSILK	QOLENNYMTFNVVKDIL	NSRFNKRNFK 1326			
Db	180	VT-MDNILSGFENEYD	IVILKPLAGVYRS	ILKQIEKNIIITFN	LNDILNSRLKRRKYFL 238			
Qy	1327	NVLESDLIPYKDLTSS	NVYVVKDPYKPLN	EKRDKFLSSNYIK	DSIDTDINFANDVLGY 1386			
Db	239	DVLESDLMQFKHLS	SNNEYIIEDSFK	LLNSEQNTLLSKY	KIKESVENDIKFAQEGISY 298			
Qy	1387	KILSEKYKSDLS	TSIKYI-----	NDKQENK	EKYPFLNNIETLYKT 1427			
Db	299	EKVLAKYKDL	ESIKVKEKEFP	SPPTPPSPAKTD	EQKESKFLPFLTNIETLYNN 358			
Qy	1428	VNDKIDLFVHLEAK	VLNYTEKSNVEV	KIKELNYLKT	IDOKLADPKKNNPVG	IADLST 1487		
Db	359	LVNKIDYDLIN	LKAKINDC	NVEKDEAHV	KITKLSDLKADDKID	ILFKNTNDFEAIKKLIN 418		
Qy	1488	DYHNHNNILF	STGWFENLAKT	IVLSNLLDGLN	GLMLNSIQHCVK	KQCPNSGCFRHL 1547		
Db	419	DDTKKMLG	LLSTGLV-QN	FPWTIISK	LIECKFQDMLN	ISQHCVKKQCPNSGCFRHL 477		
Qy	1548	DERECKCLLNY	KQEGDKCVEN	PNPTCNENNG	CGDADAKCTE	EDSGSGNGKKITCECTKPD 1607		
Db	478	DERECKCLLNY	KQEGDKCVEN	PNPTCNENNG	CGDADAKCTE	EDSGSGNGKKITCECTKPD 537		
Qy	1608	SYPLDFGIFC	SSNFLGIS	FLLLILMLYSFI	1639			
Db	538	SYPLDFGIFC	SSNFLGIS	FLLLILMLYSFI	569			
RESULT	28							
ID	Q25977	PRELIMINARY;	PRT;	569	AA.			
AC	Q25977;							
DT	01-NOV-1996	(TEMBLrel. 01, Created)						
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)						
DE	01-MAR-2002	(TEMBLrel. 20, Last annotation update)						
DE	Major merozoite surface protein (Fragment).							
GN	MSPI.							
OS	Plasmodium falciparum.							
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.							
OX	NCBI_TaxID=5833;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RP	MEDLINE=93295445; PubMed=8515786;							
RA	Jonquaites S., Tanabe K., Kanbara H.;							
RT	"Sequence conservation in the C-terminal part of the precursor to the							
RT	major merozoite surface proteins (MSPI) of Plasmodium falciparum from							
RT	field isolates."							
RL	Mol. Biochem. Parasitol. 59:95-100(1993).							
DR	EMBL; D13350; BAA02611.1;							
DR	InterPro; IPR000561; EGF-like.							
DR	Pfam; PF00008; EGF; 1							
KW	EGF-like domain; Merozoite.							
FT	NON_TER							
SQ	SEQUENCE	569	AA;	64485	MW; A26C45914AF01396 CRC64;			

	Query Match	Best Local Similarity	21.0%;	Score 11/2;	DB 5;	Length 569;
	Matches 347;	Conservative 79;	Mismatches 110;	Indels 36;	Gaps	
Qy 1101	PLKLTSESIQTEDNYASLENFVLSKLGKGLKLDNLNLEKKLSYSSGLHHLIAELKEV	1160				
Db 1	PLKLTSEVSQTEDNYANLEKFPVLSKIDGKLDNLHLGKKLFSLSGLHHLITELKEV	60				
Qy 1161	IKKNKNTGNSPSENNTDVNNALSESYYKKFLPEGTADVAVTVYSE	-----SGSDT 1206				
Db 61	IKKNKNTGNSPSENKKKNVKAALSKYENFLPE-AKVTVTYTPQDPVTPSPISVRVSGSG	119				

Qy	1207	LEQSQPKPASTHGAESN	ITTSQNVDDVDDVI	PTFGESEBDYD	DLQGVVTG	EA	1207	1366
Db	120	STKEETQPTSGSLLT	DLQVVQVQVQVQVQ	DEEDSLVLP	TFGESENDEY	DLQVV	120	179
Qy	1267	PSVDNLSKIENEYEV	LYLPLAGVYRSLK	KOLENNVFN	VNVVKD	ILNSR	1267	1326
Db	180	VT-MDNLSGFENEY	DVYLYLPLAGVYR	SLKQIEKNI	ITFNLND	ILNSRLK	180	238
Qy	1327	NVLESILPYKDLT	SSNVVQVQVQVQ	PKFPLNKKR	KDKFLSSVNY	IKOSIDT	1327	1386
Db	239	DVLESDLMQKFH	SSNEYIIE	DSFKLLN	SEQNTLLK	YKYIKESV	239	298
Qy	1387	KILSEKYS	DLSDIKYI	-----	NDKQGENE	KYLPFLNN	1387	1427
Db	299	EKLVAKYD	LESILKIV	KEEFSP	PTTTPPS	PAKTDEQ	299	358
Qy	1428	VNDKIDFVHL	SAKVLN	TYEKS	NEVYKIELN	YLKTIQ	1428	1487
Db	359	LVNKIDDY	LNLKAK	INDC	NEKDEAH	KITKLS	359	418
Qy	1488	DYNNUNLLT	FLSTG	VFNLA	KTVL	SNLLD	1488	1547
Db	419	DOTKMD	GLKLLST	GLV-QNF	NTIIS	KIEG	419	477
Qy	1548	DERECK	CLLNYKQ	EGDK	CVENP	NTC	1548	1607
Db	478	DERECK	CLLNYKQ	EGDK	CVENP	NTC	478	537
Qy	1608	SYPLFDG	IFC	SSN	FLG	ISFLIL	1608	1639
Db	538	SYPLFDG	IFC	SSN	FLG	ISFLIL	538	569

```

RESULT 29
Q25974
ID Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; Pubmed=8515786;
RT Jongwutives S., Tanabe K., Kanbara H.;
RA "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
RL EMBL; D13348; BAA02609.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER
FT 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 21.0%; Score 1768; DB 5; Length 569;
Best Local Similarity 60.5%; Pred. No. 3e-48;
Matches 346; Conservative 79; Mismatches 111; Indels 36; Gaps 5;

```

QY	1101	PKLTLSESIQTEDNYASLENFKVLUSKLEGLKDNLNLEKKLSYLSGSLHLLAEKEY	1160
		: :	
Db	.	PLKLTSEVSIQTEDNYANLEFRALSKIDGKLNDNLHLGGKKLSFLSGSLHLLTELKEY	60
QY	1161	IKNKNTGNSPENNTDVNNALESYKKLPGETDVAIVVS-	1206
Db	61	INKNKNTGNSPENNKKNYEALKSVENLFPE-AKVTTVTVPQPDVTPSPLSVRVSSGS	119

Qy	1207	LEQSOPKPKPASTHVAGBSNTITTSQNWDEVDVVIIVPIFGESEEDYDQDLGOVVTGEAVT	1266
Db	120	STKEETQPTSGSLTLEQWQVLQWDEBDSLVVLPFGESEDNDEYDQWVVTGEAIS	179
Qy	1267	PSVIDNTILSKENEYEVLYLKPLAGVYRSILKQLENNVMTFNVVVKDILNSRPNKRENFK	1326
Db	180	VT-MDNILSGFENEDYVILYKPLAGVYRSILKQIEKNITFTNLNLDILNSRLKRRYFL	238
Qy	1327	NVLESDILPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSYIKDSITDIDINFANDVLGY	1386
Db	239	DVLESDLMQFKHISSNEYIIEDSEFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISY	298
Qy	1387	KILSEKYKSDLSITKKYI-----NDKQGENEKYLPFTLNNTIETLYKT	1427
Db	299	EKVLUAKYDOLLESITKKYIKVEKEFPSPPTPPSPAKTDEQKKESKFLPFTNIETLYNN	358
Qy	1428	VNDRIDLFVHLEAKVNLTYEKSVEYKIKELNLYKTQDKLADFKKNNNFVGTADLST	1487
Db	359	LVNKIDLYLNLKAKINDCQNVKEAHVYITKLSDLKAIDDKIDLFKYTNDFEAIKKLIN	418
Qy	1488	DYNNHNLTLFTSGWVFENLAKTVLSNLDGNLOGMLNISOHQCVKKQCPNSGCFRHL	1547
Db	419	DDTKKMDLGLKLLTGLV-QNFPNTITSKLIEGKFDMLNISOHQCVKKQCPNSGCFRHL	477
Qy	1548	DERECCLLNYKQEGDKCVENPNPTCNENNGCCDADAKCTEEDSGSGKKITCECTKPD	1607
Db	478	DERECCLLNYKQEGDKCVENPNPTCNENNGCCDADAKCTEEDSGSGKKITCECTKPD	537
Qy	1608	SYPLDFGIFCSSNFILGIFLLILMLILYSFI	1639
Db	538	SYPLDFGIFCSSNFILGIFLLILMLILYSFI	569

RESULT 30

Q25979		PRELIMINARY;	PRT;	569 AA.
ID	Q25979			
AC	Q25979;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
DE	Major merozoite surface protein (Fragment).			
DE	MSPI.			
GN	Plasmodium falciparum.			
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93295445;	PubMed=8515786;		
RA	Jonglutiwes S., Tanabe K., Kanbara H.;			
RT	"Sequence conservation in the C-terminal part of the precurs-			
RT	major merozoite surface proteins (MSP) of Plasmodium falci-			
RT	p field isolates.";			
RL	Mol. Biochem. Parasitol. 59:95-100(1993).			
DR	EMBL; D13352; BAA02613.1;			
DR	InterPro; IPR000561; EGF-like.			
DR	Pfam; PF00008; EGF; 1.			
KW	EGF-like domain; Merozoite.			
FT	NON_TER			
SQ	SEQUENCE	569 AA;	64457 MW;	A2691FE56BE80396 CRC64;

Query Match 21.0%; Score 1758; DB 5; Length 569;
Best Local Similarity 60.5%; Pred. No. 3e-48;
Matches 346; Conservative 79; Mismatches 111; Indels 36; Gaps 5;

QY	1101	PKLTLSESIQEDNYASLENPKVLSKLEGLKONLNLLEKKKLSYLSGGLHHLLTAELUEV	1160
		: : : :	
Db	1	PLKLTSEVSIQEDNYANLEKFRALSKIDGLNLDNLHGKKKLSFSLSGGLHHLLITELKEV	60
QY	1161	IRKNKNTGSPENNTDVNNALSYKKFLPEGTQVATVWSE-----SGSDT	1206
Db	61	IRKNKNTGSPENNKKNYEAALKYENFLPE-AKVTVTTPPQPDVTPPSPLSVRVSGSG	119
QY	1207	LEQSQPKKPASTHVGAESENTITTSQNVDEDDVILVIPICESEEDYDDLGVVTVGEAVT	1366


```
Db 120 STKEETQIPTSGSLTTELQOVVOLQNYDEEDDSLVVLPPIFGESEDNDEYLDQVVTGEATS 179
QY 1267 PSVIDNILSKINEVEVLKPLAGVYRSLLKQLENNVMTFVNVKDIILNSRFNKRENEK 1326
Db 180 VT-MDNILSGFENEVDIYKPLAGVYRSLLKQLENNVMTFVNVKDIILNSRLKRRKRYEL 238
QY 1327 NVLESDLPYKDLTSSNVVVKPYKFLNKKRDKKFLSSNYIKDSITDITDINFANDVLGY 1386
Db 239 DVLSDLMQFKHISSEYIIESFLLNSEQNTLLSKYKIKESVENDIKFAQEGISY 298
QY 1387 KILSEKYSDDLDSIKKIYI-----NDKQGENKYLFPFLNNIETLYKT 1427
Db 299 EKVLAKYKDLLESIKKIYKEERFSSPPTTPPSPAKTDEQKESKFLPFLNIETLYNN 358
QY 1428 VNDKIDLFVIHLEAKVNLTYEKSNEVEVKIKELNYLKTIDQLADFKNNVFGIADLST 1487
Db 359 LVNKKIDYILINUKAKINDCNVEKDEAHVAKITKLSDLKAIIDDKIDLFKNTDFAIKKLIN 418
QY 1488 DYNHNNLLTKFLSTGMVFENLAKTVLSNLLDNLQMLNISQHCVKKQCPQNSGCFRHL 1547
Db 419 DDTKMDLGLKLLSTGLV-QNFPNTIISKLEGFQDMLNISQHCVKKQCPQNSGCFRHL 477
QY 1548 DERECKLLNTYKQEGDKCVENPNTCENNNGCGDADAKCTEEDSGSKKKTCTCKPD 1607
Db 478 DERECKLLNTYKQEGDKCVENPNTCENNNGCGDADAKCTEEDSGSKKKTCTCKPD 537
QY 1608 SYPLFDGIFCSSNFGISFLILMLILYSFI 1639
Db 538 SYPLFDGIFCSSNFGISFLILMLILYSFI 569
```

```
RESULT 31
Q25983
ID Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSP1) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13355; BAA02616.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64491 MW; 57A6B62EF72CB885 CRC64;
```

Query Match 21.0%; Score 1767; DB 5; Length 569;
Best Local Similarity 60.5%; Pred. No. 3.2e-48;
Matches 346; Conservative 78; Mismatches 112; Indels 36; Gaps 5;

```
QY 1101 PKLTISESIQTDENVASLENFKVLSKLGKLDNLNLEKKKLSYLSGLHHLIAELKEV 1160
Db 1 PKLTISESIQTDENVANLEKFRVLSKIDGKLDNLNHLGKKLSFLSSGLHHLITELKEV 60
QY 1161 IKKNYTGNSPSENNTDVNNALLESYKKFLPECTDVATVYSE-----SGSDT 1206
Db 61 IKKNYTGNSPENNNKKNVNEALKSYENFLPE-AKVTVTVPQDPVTPSPLSRVSGSG 119
QY 1207 LEQSQPKPASTHVGAESENITTSQNVDDVDVIIVPIFGESEEDYDLQGVVTGEAVT 1266
```

```
Db 120 STKEETQIPTSGSLTTELQOVVOLQNYDEEDDSLVVLPPIFGESEDNDEYLDQVVTGEATS 179
QY 1267 PSVIDNILSKINEVEVLKPLAGVYRSLLKQLENNVMTFVNVKDIILNSRFNKRENEK 1326
Db 180 VT-MDNILSGFENEVDIYKPLAGVYRSLLKQLENNVMTFVNVKDIILNSRLKRRKRYEL 238
QY 1327 NVLESDLPYKDLTSSNVVVKPYKFLNKKRDKKFLSSNYIKDSITDITDINFANDVLGY 1386
Db 239 DVLSDLMQFKHISSEYIIESFLLNSEQNTLLSKYKIKESVENDIKFAQEGISY 298
QY 1387 KILSEKYSDDLDSIKKIYI-----NDKQGENKYLFPFLNNIETLYKT 1427
Db 299 EKVLAKYKDLLESIKKIYKEERFSSPPTTPPSPAKTDEQKESKFLPFLNIETLYNN 358
QY 1428 VNDKIDLFVIHLEAKVNLTYEKSNEVEVKIKELNYLKTIDQLADFKNNVFGIADLST 1487
Db 359 LVNKKIDYILINUKAKINDCNVEKDEAHVAKITKLSDLKAIIDDKIDLFKNTDFAIKKLIN 418
QY 1488 DYNHNNLLTKFLSTGMVFENLAKTVLSNLLDNLQMLNISQHCVKKQCPQNSGCFRHL 1547
Db 419 DDTKMDLGLKLLSTGLV-QNFPNTIISKLEGFQDMLNISQHCVKKQCPQNSGCFRHL 477
QY 1548 DERECKLLNTYKQEGDKCVENPNTCENNNGCGDADAKCTEEDSGSKKKTCTCKPD 1607
Db 478 DERECKLLNTYKQEGDKCVENPNTCENNNGCGDADAKCTEEDSGSKKKTCTCKPD 537
QY 1608 SYPLFDGIFCSSNFGISFLILMLILYSFI 1639
Db 538 SYPLFDGIFCSSNFGISFLILMLILYSFI 569
```

```
RESULT 32
Q25967
ID Q25967 PRELIMINARY; PRT; 569 AA.
AC Q25967;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Major merozoite surface protein (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSP1) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13344; BAA02605.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF; 1.
KW EGF-like domain; Merozoite.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64435 MW; BE600D346E51304E CRC64;
```

Query Match 21.0%; Score 1765; DB 5; Length 569;
Best Local Similarity 60.5%; Pred. No. 3.7e-48;
Matches 346; Conservative 79; Mismatches 111; Indels 36; Gaps 5;

```
QY 1101 PKLTISESIQTDENVASLENFKVLSKLGKLDNLNLEKKKLSYLSGLHHLIAELKEV 1160
Db 1 PKLTISESIQTDENVANLEKFRVLSKIDGKLDNLNHLGKKLSFLSSGLHHLITELKEV 60
QY 1161 IKKNYTGNSPSENNTDVNNALLESYKKFLPECTDVATVYSE-----SGSDT 1206
Db 61 IKKNYTGNSPENNNKKNVNEALKSYENFLPE-AKVTVTVPQDPVTPSPLSRVSGSG 119
QY 1207 LEQSQPKPASTHVGAESENITTSQNVDDVDVIIVPIFGESEEDYDLQGVVTGEAVT 1266
Db 120 STKEETQIPTSGSLTTELQOVVOLQNYDEEDDSLVVLPPIFGESEDNDEYLDQVVTGEATS 179
```


Qy	1367	PSVIDNTLSKIENEYEVLYLKPLAGVYRSKQLENNVMWTFNVVKDILNSRFNKRENFK	1326
		: : : ::: : : : : :: : : :	
Db	180	VT-MDNLTSGFENEYDVYLKPLAGVYRSKKQIEKNIIITFNLDNLNLRKKRYFL	238
Qy	1327	NVLSDLPYKOLTSSTNYVVKDPYFLNKEKRDKFLSSVNYIKOSIDTDINFANDVLCYY	1386
		: :: : : : : : : : : : : : : : : : : : : : :	
Db	239	DVLESDLMQGFHSHSNEYIIESFKLLNSEQNRTLKSKYKYIKESVENDIKFAQEGISYY	298
Qy	1387	KILSEYKSDLDISKIYY-----NDKOGENKYLPLFLNNIETLYK	1426
		: : : :: : : : : : : : : : : : : : :	
Db	299	EKLAKYKDDLESIKVIKEEKPPSPPTTTPPSPAKTDEQKESKELPFLTNIETLYN	358
Qy	1427	TVDNKDILFVIHLAEKVLNYYTEKS NVFYK I E L N Y L K T I O Q D K A D F P K N N N V F G C I A D L S	1486
		: : :: : : : : : : : : : : : : : : : : : :	
Db	359	NLVNKIDDYLIINLKAINDCNVEKDBAHYKITKLSDLKAIDDKIDLFKNTNDFEAIKKLI	418
Qy	1487	TDYNHNHLLPKFLSTGMVFENIAKTVLSNLLDGNLQGMLNISHOCHVKQKCQPONS GC FRH	1546
		: : : : : : : : : : : : : : : : : : : : :	
Db	419	NDDTKDMGLCKLLSTGLV-QNEPNTTIISKLGKFDMDMLNISOHCYVKVKQCPE NSGC FRH	477
Qy	1547	LDERECKCLLNYKQEGDKCVENPNPTCNRNNGGCDADAKCTEEDSGSNKKKICTECTPK	1606
		: :: : : : : : : : :	
Db	478	LDERECKCLLNYKQEGDKCVENPNPTCNRNNGGCDADATCTEEDSGSRKKICTECTPK	537
Qy	1607	DSYPLFDGIFCSSNPLGISFLILMLIYSFI 1639	
		: :: : : : : : : :	
Db	538	DSYPLFDGIFCSSNPLGISFLILMLIYSFI 570	
RESULT 34			
ID	Q25982	PRELIMINARY;	PRT; 569 AA.
AC	Q25982;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DR	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Major merozoite surface protein (Fragment).		
GN	MSP1.		
OS	Plasmodium falciparum..		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
NCBI_Taxid	-5833;		

	Query Match	20.1%;	Score 1753;	DB 3;	Length 309;
	Best Local Similarity	60.1%;	Pred. No. 8.8e-48;		
	Matches 344;	Conservative 80;	Mismatches 112;	Indels 36;	Gaps
Qy	1101	PLKLTSESTQTDENVASLENKVLKSGCKLKDNLNLEKKKLSVLSGLHLLHLLAEKV	1160		
Db	1	PLKLTSEVSIQTDEYNTANLEKFLVLSKIDGKLNLDLHLCKKLSPLSSGLHLLITELKEV	60		
Qy	1161	IKKNKNTGNSPSNNNTDVNNALLESYKKFLPEGCTDVATVYSE-----SGSDT	1206		
Db	61	IKKNKNTGNSPSNNKVNKALESYENFLPE-AKVTYVTPQDPVTPSPLSRVVSGSG	119		
Qy	1207	LEGSOPKKPASTHWGAESENTITTSQNVDDVDDVIIVPIFGESEDYDLDGOWVTGEAVT	1266		
Db	120	STREETQIPVSGSLTELOQVQVLQNVDEEDSLVVLPIFGESDENDEYLOQVVTGEALS	179		
Qy	1267	PSVIDNLSIKIENEYEVLYLKLPLAGVYRSRLKQLENNVMTFNVNVKDIUNSRFNKRENF	1326		


```
Db 120 STKEETQIPTSGLLTELQOVVOLQYDEEDSLVVLPIPFGESEDNEDYLDQVVTGEAIS 179
QY 1267 PSVIDNLSKIENEYEVLYLKLPLAGVYRSUKKOLENNVMTFNVVKDILNSRFNKRENF 1326
Db 180 VT-MDNILSGFENEYDVLYLKLPLAGVYRSUKKOLENNVMTFNVVKDILNSRLKRRYFL 238
QY 1327 NVLESOLIPYKOLTSYVVKVQPYKFLNKEKRDKFLSSYIKDSITDINFANDVLGY 1386
Db 239 DVLESOLMQFKHTSSNEYIIFEDSKLLNSQKNTLLSKYIKESVENDIKFAQEGISY 298
QY 1387 KILSEYKSDLDISIKYI-----NDKQGENEKYLPFLNNIETLYKT 1427
Db 299 EKVLAKYKDDLESIKKVIKEEFPPPTPPSPAKTDEQKESKELFPLTNIEIYLYNN 358
QY 1428 VNDKIDLFVHLAKVLTNYETKSNVEVKIKELNYLKTIDQKLADFKNNNFVGIADLST 1487
Db 359 LVNKIDDYLINKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKLLIN 418
QY 1488 DYNHNNLLTKFLSTGMVFENLAKTVISNLDGNLQGLMNTSHQCQVKKQCPQNSGCFRHL 1547
Db 419 DDTKKDMLGKLLSTGLV-QNFPNTLIISKLTIEGKFQDMLNISOHQCVKKQCPENSGCFRHL 477
QY 1548 DERECKCLLNYKQEGDKCVENPNTCENNNGGCDADAKTEEDSGSNGKKITCECTKPD 1607
Db 478 DERECKCLLNYKQEGDKCVENPNTCENNNGGCDADAKTEEDSGSRRKKITCECTKPD 537
QY 1608 SYPLFDGIFCSSNFIIGISFLILMLILYSFI 1639
Db 538 SYPLFDGIFCSSNFIIGISFLILMLILYSFI 569

RESULT 39
ID Q25865 PRELIMINARY; PRT; 336 AA.
AC Q25865;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91304517; PubMed=1852173;
RA Kerr P.J., Ranford-Cartwright L.C., Walliker D.;
RT "Proof of intragenic recombination in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 46:185-187(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95107347; PubMed=7808474;
RA Kerr P.J., Ranford-Cartwright L.C., Walliker D.;
RT "Proof of intragenic recombination in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:241-248(1994).
DR EMBL; X52962; CAA37135.1; -.
FT NON_TER 1
FT NON_TER 336
SQ SEQUENCE 336 AA; 37906 MW; 3094DOC5102B73DE CRC64;

Query Match 20.1%; Score 1696; DB 5; Length 336;
Best Local Similarity 96.0%; Pred. No. 3.1e-46;
Matches 334; Conservative 1; Mismatches 1; Indels 12; Gaps 1;

QY 21 THESYOELVKLEALEDAVLTYSLFQKEKMYLNCTGTAVTTSTPGSGSVASGSGG 80
Db 1 THESYOELVKLEALEDAVLTYSLFQKEKMYLNCTGTAVTTSTPGSGSVASGSGG --- 57
QY 81 SVASGGSVASGGSVASGGSNSRRTNPSONSSDASKYADLKHVRNYLLTITK 140
Db 58 -----SGGSVASGGSVASGGSNSRRTNPSONSSDASKYADLKHVRNYLLTITK 108
QY 141 ELKYPOLFDTLNMHLTLCDNIHGFKYLDGYEEINELLYKLNFDLLRAKLNDCVANDY 200
```

```
Db 109 ELKYPOLFDTLNMHLTLCDNIHGFKYLDGYEEINELLYKLNFDLLRAKLNDCVANDY 168
QY 201 CQIPPNLKITRANEDLVLLKLVFGYRKPDLNIDKQVGMEDYIKNNKKTININELIEESK 260
Db 169 CQIPPNLKITRANEDLVLLKLVFGYRKPDLNIDKQVGMEDYIKNNKKTININELIEESK 228
QY 261 KTDKNKNATKEBEKKKLYQAOYDLSYNNKQLEEAHNLISVLEKRIDTFLAKNENIKELLD 320
Db 229 KTDKNKNATKEBEKKKLYQAOYDLSYNNKQLEEAHNLISVLEKRIDTFLAKNENIKELLD 288
QY 321 KINEIKNPPANGNPNTLLDKNKKIEEHEKEIKEIAKTIKFNIDSL 368
Db 289 KINEIKNPPANGNPNTLLDKNKKIEEHEKEIKEIAKTIKFNIDSL 336

RESULT 40
Q9TZU5
ID Q9TZU5 PRELIMINARY; PRT; 356 AA.
AC Q9TZU5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HN2;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061143; AAC69742.1; -.
FT NON_TER 356
FT NON_TER 356
SQ SEQUENCE 356 AA; 39756 MW; 35042CC13AB890E6 CRC64;

Query Match 19.8%; Score 1670; DB 5; Length 356;
Best Local Similarity 91.6%; Pred. No. 2.1e-45;
Matches 329; Conservative 10; Mismatches 14; Indels 6; Gaps 2;

QY 1 MKIIFLCFLFFIINTQCVTHESYOELVKKLEALEDAVLTYSLFQKEKMYLNCTGT 60
Db 1 MKIIFLCFLFFIINTQCVTHESYOELVKKLEALEDAVLTYSLFQKEKMYLNCTGT 60
QY 61 AVTTSTPGSK---GSVASGGSVASGGSVASGGSVASGGSNSRRTNPDSNS 117
Db 61 AVTTSTPGSKGSGSVASGGSVASGGSVASGGSVA---SVASGGSNSRRTNPDSNS 117
QY 118 SDSDAKSYADLKHVRNYLLITIKELYPOLFDTLNMHLTLCDNIHGFKYLDGYEEINEL 177
Db 118 SDSDAKSYADLKHVRNYLLITIKELYPOLFDTLNMHLTLCDNIHGFKYLDGYEEINEL 177
QY 178 LYKLNIFYDILLRAKLNDCVANDYQIPFNLIKIRANELDVLKLVFGYRKPDLNIDKQVNGK 237
Db 178 LYKLNIFYDILLRAKLNDCVANDYQIPFNLIKIRANELDVLKLVFGYRKPDLNIDKQVNGK 237
QY 238 MEDYIKNNKKTININELIEESKTTDKNNKATKEBEKKKLYQAOYDLSYNNKQLEEAHN 297
Db 238 MEDYIKNNKKTININELIEESKTTDKNNKATKEBEKKKLYQAOYDLSYNNKQLEEAHN 297
QY 298 LISVLEKRIDTFLAKNENIKELLDKINEIKNPPANGNPNTLLDKNKKIEEHEKEIKE 356
Db 298 LISVLEKRIDTFLAKNENIKELLDKINEIKNPPANGNPNTLLDKNKKIEEHEKEIKE 356

RESULT 41
Q9TZU2
```



```
ID O9TZ02 PRELIMINARY; PRT; 350 AA.
AC O9TZ02;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HN6.24;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061146; AAC69745.1; -.
FT NON_TER 350
SQ SEQUENCE 350 AA; 39343 MW; ALE3C3AA6AF48FC9 CRC64;

Query Match 19.6%; Score 1649; DB 5; Length 350;
Best Local Similarity 91.0%; Pred. No. 9.4e-45;
Matches 324; Conservative 11; Mismatches 15; Indels 6; Gaps 2;

QY 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEKMWLNEGTS 60
Db 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEKMWLNEGTS 60
QY 61 AVTTSTPGSGVTSVSGSGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSG 120
Db 61 AVTTSTPGSGSVT---SGGSVTSGG---SGGSVASVASGSGSVASGSGSVASG 114
QY 121 DAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLIDGYEIN 180
Db 115 DAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLIDGYEIN 174
QY 181 LNFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKN 240
Db 175 LNFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPLDNIKN 234
QY 241 YIKKNTTIENINELIEESKTTIDKNKNATKEEKKLYQAOYDLSYNNKOLEEHN 300
Db 235 YIKKNTTIENINELIEESKTTIDQKNADNEEGKKLYQAOYDLSYNNKOLEEHN 294
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTNPNTLLDKNKKIEHEKEIK 356
Db 295 VLEKRIDTLKKNENIKKLDKIDEIKNPPANSNGTNPNTLLDKNKKIEHEKEIK 350

RESULT 42
Q9TZ07 PRELIMINARY; PRT; 357 AA.
ID Q9TZ07;
AC Q9TZ07;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA125;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
```

```
Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061141; AAC69740.1; -.
FT NON_TER 357
SQ SEQUENCE 357 AA; 39850 MW; AE9513F521FA9ED5 CRC64;

Query Match 19.4%; Score 1636.5; DB 5; Length 357;
Best Local Similarity 90.6%; Pred. No. 2.4e-44;
Matches 326; Conservative 11; Mismatches 16; Indels 7; Gaps 3;

QY 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEKMWLN 60
Db 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEKMWLN 60
QY 61 AVTTSTPG---SKGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSVAS 117
Db 61 AVTTSTPGSVASGSGSVASGSGSVASGSG---SGGSVASGSGSVASGSGSVAS 117
QY 118 SDSAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLID 177
Db 118 SDSAKSYADLKHVRNLYLTIKELYPQLFDLTNHLMLTCLDNIHGFYKYLID 177
QY 178 LYKLNIFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPL 237
Db 178 LYKLNIFYDILLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKPL 237
QY 238 MEDYIKKNTTIENINELIEESKTTIDKNKNATKEEKKLYQAOYDLSYNNK 297
Db 238 MEDYIKKNTTIENINELIEESKTTIDQKNADNEEGKKLYQAOYDLSYNNK 297
QY 298 LISVLEKRIDTLKKNENIKELLDKINEIK-NPPANSNGTNPNTLLDKNKKIE 356
Db 298 LISVLEKRIDTLKKNENIKKLEDDIKTDAEKLTTGSKPNTLLDKNKKIEHE 357

RESULT 43
Q9TZ08 PRELIMINARY; PRT; 357 AA.
ID Q9TZ08;
AC Q9TZ08;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA12;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061140; AAC69739.1; -.
FT NON_TER 357
SQ SEQUENCE 357 AA; 39850 MW; 498FCD8E2DF1A1D CRC64;

Query Match 19.4%; Score 1634.5; DB 5; Length 357;
Best Local Similarity 90.3%; Pred. No. 2.7e-44;
Matches 325; Conservative 13; Mismatches 15; Indels 7; Gaps 3;

QY 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEKMWLN 60
Db 1 MKIIFLCSEFLFIINTQCVTHESYOELVKKLEALEDAVLGTSLFQKEKMWLN 60
QY 61 AVTTSTPG---SKGSVASGSGSVASGSGSVASGSGSVASGSGSVASGSGSVAS 117
Db 61 AVTTSTPGSVASGSGSVASGSGSVASGSG---SGGSVASGSGSVASGSGSVAS 117
```



```
Qy 118 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCNDINHGFYKYLIDGYEEINEL 177
Db 118 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCNDINHGFYKYLIDGYEEINEL 177
Qy 178 LYKLNFFDLLRAKLVNDVANDYCOIPENLKIIRANELDVLKLVFGYKRPDLNIDKNVYK 237
Db 178 LYKLNFFDLLRAKLVNDVANDYCOIPENLKIIRANELDVLKLVFGYKRPDLNIDKNVYK 237
Qy 238 MEDYIKKKNTTIANINELIEESKKTIDKNKATKEEKKLYQAOYDLSIYNKQLEEAHN 297
Db 238 MEDYIKKKNTTIANINELIEESKKTIDKNKADNEEGKKLYQAOYDLSIYNKQLEEAHN 297
Qy 298 LISVLEKRDITLKKNENIKELLDKINEIK-NPPPPANSNGTPTNTLDDKNKKIEEHEKEIKE 356
Db 298 LISVLEKRDITLKKNENIKELLDKINEIK-NPPPPANSNGTPTNTLDDKNKKIEEHEKEIKE 357
RESULT 44
Q9TZV2 ID Q9TZV2 PRELIMINARY; PRT; 360 AA.
AC Q9TZV2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFA9.2;
RA JIANG G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061136; AAC69735.1; -.
FT NON_TER 360
SQ SEQUENCE 360 AA; 40170 MW; 224E3CD345BD2F7A CRC64;
Query Match 19.3%; Score 1625.5; DB 5; Length 360;
Best Local Similarity 90.0%; Pred. No. 5.2e-44;
Matches 325; Conservative 12; Mismatches 19; Indels 5; Gaps 3;
Qy 1 MKIIFFLCSEFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Db 1 MKIIFFLCSEFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Qy 61 AVTTSTPGSKVSGSGGSA---SGSVASGSGSVASGSGSGSGNSRRTPNSDS 117
Db 61 AVTTSTPGSKVSGSGGSA---SGSVASGSGSVASGSGSGSGNSRRTPNSDS 120
Qy 118 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCNDINHGFYKYLIDGYEEINEL 177
Db 121 SDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCNDINHGFYKYLIDGYEEINEL 180
Qy 178 LYKLNFFDLLRAKLVNDVANDYCOIPENLKIIRANELDVLKLVFGYKRPDLNIDKNVYK 237
Db 181 LYKLNFFDLLRAKLVNDVANDYCOIPENLKIIRANELDVLKLVFGYKRPDLNIDKNVYK 240
Qy 238 MEDYIKKKNTTIANINELIEESKKTIDKNKATKEEKKLYQAOYDLSIYNKQLEEAHN 297
Db 241 MEDYIKKKNTTIANINELIEESKKTIDKNKADNEEGKKLYQAOYDLSIYNKQLEEAHN 300
Qy 298 LISVLEKRDITLKKNENIKELLDKINEIK-NPPPPANSNGTPTNTLDDKNKKIEEHEKEIKE 356
Db 301 LISVLEKRDITLKK-ENIKKLELDIKTDAEKPPTTESKPTNTLDDKNKKIEEHEKEIKE 359
Qy 357 I 357
```

```
Db 360 I 360
RESULT 45
Q9TZU9 ID Q9TZU9 PRELIMINARY; PRT; 363 AA.
AC Q9TZU9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFA11;
RA JIANG G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061139; AAC69738.1; -.
FT NON_TER 363
SQ SEQUENCE 363 AA; 40398 MW; C25F38CB636941B7 CRC64;
Query Match 19.3%; Score 1625.5; DB 5; Length 363;
Best Local Similarity 89.0%; Pred. No. 5.3e-44;
Matches 323; Conservative 15; Mismatches 18; Indels 7; Gaps 3;
Qy 1 MKIIFFLCSEFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Db 1 MKIIFFLCSEFFIINTOCVTHESYQELVKKLEALEDVLTGYSLFQKEKMWLNEGTSGT 60
Qy 61 AVTTSTPGSKVSGSGGSA---SGSVASGSGSVASGSGSGSGNSRRTPNS 114
Db 61 AVTTSTPGSKVSGSGGSA---SGSVASGSGSVASGSGSGSGNSRRTPNS 120
Qy 115 DNSSDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCNDINHGFYKYLIDGYEEI 174
Db 121 DNSSDSAKSYADLKHVRVNYLLTIKELKYPQLFDLTNHMLTLCNDINHGFYKYLIDGYEEI 180
Qy 175 NELLYKLNFFDLLRAKLVNDVANDYCOIPENLKIIRANELDVLKLVFGYKRPDLNIDKN 234
Db 181 NELLYKLNFFDLLRAKLVNDVANDYCOIPENLKIIRANELDVLKLVFGYKRPDLNIDKN 240
Qy 235 VGMEDYIKKKNTTIANINELIEESKKTIDKNKATKEEKKLYQAOYDLSIYNKQLEE 294
Db 241 VGMEDYIKKKNTTIANINELIEESKKTIDKNKADNEEGKKLYQAOYDLSIYNKQLEE 300
Qy 295 AHNLSVLEKRDITLKKNENIKELLDKINEIK-NPPPPANSNGTPTNTLDDKNKKIEEHEKE 353
Db 301 AHNLSVLEKRDITLKKNENIKELLDKINEIK-NPPPPANSNGTPTNTLDDKNKKIEEHEKE 360
Qy 354 IKE 356
Db 361 IKE 363
RESULT 46
Q9TZU1 ID Q9TZU1 PRELIMINARY; PRT; 344 AA.
AC Q9TZU1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```



```
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hn6.27;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RT Pluschke G.;
RA "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061147; AAC69746.1; -.
FT NON_TER 344
SQ SEQUENCE 344 AA; 38767 MW; 1225C0E1683007DF CRC64;

Query Match 19.3%; Score 1622; DB 5; Length 344;
Best Local Similarity 89.6%; Pred. No. 6.4e-44;
Matches 319; Conservative 11; Mismatches 14; Indels 12; Gaps 2;

QY 1 MKIIFFLCFLFFIINTQCVTHESYOELVKLEALEDAVLTYGSLFQKEKMWLNESGTS 60
Db 1 MKIIFFLCFLFFIINTQCVTHESYOELVKLEALEDAVLTYGSLFQKEKMWLNESGTS 60
QY 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNPDSN 120
Db 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNPDSN 120
QY 121 DAKSYADLKHVRVNYLLTIKELKYPOLFDLTNHLMTLCNDIHGFKYLDIGYEINELLYK 180
Db 121 DAKSYADLKHVRVNYLLTIKELKYPOLFDLTNHLMTLCNDIHGFKYLDIGYEINELLYK 171
QY 181 LNFYFDLLRAKLVNDVANDYCOIPFNLKIRANDELVDLKLKLVFGYRKPDLNKNVGMED 240
Db 181 LNFYFDLLRAKLVNDVANDYCOIPFNLKIRANDELVDLKLKLVFGYRKPDLNKNVGMED 231
QY 241 YKKNNKTTINELIESEKKTIDKNKNATKEBEKKLYQAOYDLSIYNKQLEEAHNLIS 300
Db 241 YKKNNKTTINELIESEKKTIDKNKNATKEBEKKLYQAOYDLSIYNKQLEEAHNLIS 291
QY 301 VLEKRTDTLKKENIKELDKINEIKNPPANSNGTNTLLDKNKKIEBEKEIKEIAKT 360
Db 301 VLEKRTDTLKKENIKELDKINEIKNPPANSNGTNTLLDKNKKIEBEKEIKEIAKT 351
QY 361 IKFNIDSLETFD 371
Db 361 IKFNIDSLETFD 362

RESULT 48
Q9TZV1
ID Q9TZV1 PRELIMINARY; PRT; 365 AA.
AC Q9TZV1;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA9.16;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RA "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061137; AAC69736.1; -.
FT NON_TER 365
SQ SEQUENCE 365 AA; 40455 MW; E68D36520F4F4AE0 CRC64;

Query Match 18.9%; Score 1595; DB 5; Length 365;
Best Local Similarity 86.8%; Pred. No. 4.7e-43;
Matches 317; Conservative 16; Mismatches 24; Indels 8; Gaps 3;

QY 1 MKIIFFLCFLFFIINTQCVTHESYOELVKLEALEDAVLTYGSLFQKEKMWLNESGTS 60
Db 1 MKIIFFLCFLFFIINTQCVTHESYOELVKLEALEDAVLTYGSLFQKEKMWLNESGTS 60
QY 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNP 114
Db 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNP 120
QY 115 DNSSSDAKSYADLKHVRVNYLLTIKELKYPOLFDLTNHLMTLCNDIHGFKYLDIGYEEI 174
Db 115 DNSSSDAKSYADLKHVRVNYLLTIKELKYPOLFDLTNHLMTLCNDIHGFKYLDIGYEEI 180
QY 175 NELLKYNFYFDLLRAKLVNDVANDYCOIPFNLKIRANDELVDLKLKLVFGYRKPDLNKN 234
Db 175 NELLKYNFYFDLLRAKLVNDVANDYCOIPFNLKIRANDELVDLKLKLVFGYRKPDLNKN 234
```

```
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hn6.27;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RT Pluschke G.;
RA "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061147; AAC69746.1; -.
FT NON_TER 344
SQ SEQUENCE 344 AA; 38767 MW; 1225C0E1683007DF CRC64;

Query Match 19.3%; Score 1622; DB 5; Length 344;
Best Local Similarity 89.6%; Pred. No. 6.4e-44;
Matches 319; Conservative 11; Mismatches 14; Indels 12; Gaps 2;

QY 1 MKIIFFLCFLFFIINTQCVTHESYOELVKLEALEDAVLTYGSLFQKEKMWLNESGTS 60
Db 1 MKIIFFLCFLFFIINTQCVTHESYOELVKLEALEDAVLTYGSLFQKEKMWLNESGTS 60
QY 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNPDSN 120
Db 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGSNRRTNPDSN 108
QY 121 DAKSYADLKHVRVNYLLTIKELKYPOLFDLTNHLMTLCNDIHGFKYLDIGYEINELLYK 180
Db 121 DAKSYADLKHVRVNYLLTIKELKYPOLFDLTNHLMTLCNDIHGFKYLDIGYEINELLYK 168
QY 181 LNFYFDLLRAKLVNDVANDYCOIPFNLKIRANDELVDLKLKLVFGYRKPDLNKNVGMED 240
Db 181 LNFYFDLLRAKLVNDVANDYCOIPFNLKIRANDELVDLKLKLVFGYRKPDLNKNVGMED 228
QY 241 YKKNNKTTINELIESEKKTIDKNKNATKEBEKKLYQAOYDLSIYNKQLEEAHNLIS 300
Db 241 YKKNNKTTINELIESEKKTIDKNKNATKEBEKKLYQAOYDLSIYNKQLEEAHNLIS 288
QY 301 VLEKRTDTLKKENIKELDKINEIKNPPANSNGTNTLLDKNKKIEBEKEIKE 356
Db 301 VLEKRTDTLKKENIKELDKINEIKNPPANSNGTNTLLDKNKKIEBEKEIKE 344

RESULT 47
Q9TZV3
ID Q9TZV3 PRELIMINARY; PRT; 362 AA.
AC Q9TZV3;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFA9.16;
RX MEDLINE=20106724; PubMed=10643908;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RA "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL; AF061135; AAC69734.1; -.
FT NON_TER 362
SQ SEQUENCE 362 AA; 41401 MW; A3129A0C4A1AC600 CRC64;

Query Match 19.2%; Score 1616.5; DB 5; Length 362;
Best Local Similarity 86.8%; Pred. No. 1e-43;
Matches 322; Conservative 11; Mismatches 29; Indels 9; Gaps 2;
```



```
Db 181 NELLKLYNEFYDRLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRPLDNKDN 240
QY 235 VGMEDYIKNNKTTIENINELIEESKTTIDKNKNATKEEKKLYQAQYDLSIYNKQLEE 294
Db 241 VGMEDYIKNNKTTIENINELIEESKTTIDQKNADNEEGKKLYQAQYDLSIYNKQLEE 300
QY 295 AHNLSVLEKRDITLKKNNENIKELDKINEIK-NPPANSGMTPTNTLLD-KNKKTEEHKE 352
Db 301 AHNLSVLEKRDITLKKNNENIKELDKINEIK-NPPANSGMTPTNTLLD-KNKKTEEHKE 360
QY 353 EIKEI 357
Db 361 KIKEI 365

RESULT 49
Q9TZU4
ID Q9TZU4 PRELIMINARY; PRT; 351 AA.
AC Q9TZU4;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HN3;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL: AF061144; AAC69744.1; -.
FT NON_TER 351
SQ SEQUENCE 351 AA; 39465 MW; 06481719E004278D CRC64;

Query Match 18.8%; Score 1585.5; DB 5; Length 351;
Best Local Similarity 88.2%; Pred. No. 9e-43;
Matches 315; Conservative 16; Mismatches 19; Indels 7; Gaps 3;

QY 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
Db 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
QY 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVS 120
Db 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVS 114
QY 121 DAKSYADLKHRVRYNLLTIKELYPQLFDLTNHMLTLCNDINHGFYKLYIDGYEEINELLYK 180
Db 115 DAKSYADLKHRVRYNLLTIKELYPQLFDLTNHMLTLCNDINHGFYKLYIDGYEEINELLYK 174
QY 181 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRPLDNKDNVCKMED 240
Db 175 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRPLDNKDNVCKMED 234
QY 241 YIKKNTTIENINELIEESKTTIDKNKNATKEEKKLYQAQYDLSIYNKQLEEANLIS 300
Db 235 YIKKNTTIENINELIEESKTTIDQKNADNEEGKKLYQAQYDLSIYNKQLEEANLIS 294
QY 301 VLEKRDITLKKNNENIKELDKINEIK-NPPANSGMTPTNTLLDKNKKIEEHKEIKE 356
Db 295 VLEKRDITLKKNNENIKELDKINEIK-NPPANSGMTPTNTLLDKNKKIEEHKEIKE 351

RESULT 51
Q9TZV0
ID Q9TZV0 PRELIMINARY; PRT; 360 AA.
AC Q9TZV0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFA10;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
```

```
ID Q9TZU3 PRELIMINARY; PRT; 351 AA.
AC Q9TZU3;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HN3;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
RT Tanzania.";
RL Acta Trop. 74:51-61(2000).
DR EMBL: AF061145; AAC69744.1; -.
FT NON_TER 351
SQ SEQUENCE 351 AA; 39465 MW; 06481719E004278D CRC64;

Query Match 18.8%; Score 1585.5; DB 5; Length 351;
Best Local Similarity 88.2%; Pred. No. 9e-43;
Matches 315; Conservative 16; Mismatches 19; Indels 7; Gaps 3;

QY 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
Db 1 MKIIFLCSFLFFIINTQCVTTHESYQELVKLEALEDAVLTCYSLFQKEKMWLNESGTS 60
QY 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVS 120
Db 61 AVTTSTPGSKGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVSAGSGGVS 114
QY 121 DAKSYADLKHRVRYNLLTIKELYPQLFDLTNHMLTLCNDINHGFYKLYIDGYEEINELLYK 180
Db 115 DAKSYADLKHRVRYNLLTIKELYPQLFDLTNHMLTLCNDINHGFYKLYIDGYEEINELLYK 174
QY 181 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRPLDNKDNVCKMED 240
Db 175 LNFYDILLRAKLVNDVANDYQIPFNKIRANELDVLKLVFGYRPLDNKDNVCKMED 234
QY 241 YIKKNTTIENINELIEESKTTIDKNKNATKEEKKLYQAQYDLSIYNKQLEEANLIS 300
Db 235 YIKKNTTIENINELIEESKTTIDQKNADNEEGKKLYQAQYDLSIYNKQLEEANLIS 294
QY 301 VLEKRDITLKKNNENIKELDKINEIK-NPPANSGMTPTNTLLDKNKKIEEHKEIKE 356
Db 295 VLEKRDITLKKNNENIKELDKINEIK-NPPANSGMTPTNTLLDKNKKIEEHKEIKE 351

RESULT 51
Q9TZV0
ID Q9TZV0 PRELIMINARY; PRT; 360 AA.
AC Q9TZV0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Merozoite surface protein 1 (Fragment).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFA10;
RA Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,
RA Pluschke G.;
RT "Sequence diversity of the merozoite surface protein 1 of Plasmodium
RT falciparum in clinical isolates from the Kilombero District,
```


[illegible]

QY	117	SSDSKASYADLKURVRNYLLTTIKELKYPQLFDLTNHLMTLCDNHGFVKYLDGYEINE	178
Db <td>114</td> <td>ASDSKASYADLKURVRNYLLTTIKELKYPQLFDLTNHLMTLCDNHGFVKYLDGYEINE</td> <td>173</td>	114	ASDSKASYADLKURVRNYLLTTIKELKYPQLFDLTNHLMTLCDNHGFVKYLDGYEINE	173
QY <td>177</td> <td>LLYLNFYDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKLPLDNKDNVG</td> <td>236</td>	177	LLYLNFYDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKLPLDNKDNVG	236
Db <td>174</td> <td>LLYLNFYDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKLPLDNKDNVG</td> <td>233</td>	174	LLYLNFYDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKLPLDNKDNVG	233
QY <td>237</td> <td>KWEDIYKKNKTTIENINELIEESKTTIDKNNATKEEKKKLYOAOYDLSIYNKQLEEAH</td> <td>296</td>	237	KWEDIYKKNKTTIENINELIEESKTTIDKNNATKEEKKKLYOAOYDLSIYNKQLEEAH	296
Db <td>234</td> <td>KWEDIYKKNKTTIENINELIEESKTTIDKNNATKEEKKKLYOAOYDLSIYNKQLEEAH</td> <td>293</td>	234	KWEDIYKKNKTTIENINELIEESKTTIDKNNATKEEKKKLYOAOYDLSIYNKQLEEAH	293
QY <td>297</td> <td>NLISVLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKE</td> <td>356</td>	297	NLISVLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKE	356
Db <td>294</td> <td>NLISVLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKE</td> <td>353</td>	294	NLISVLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKE	353
RESULT 53			
Q9TZV6		PRELIMINARY; PRT; 347 AA.	
AC	09TZV6		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	Merizoite surface protein 1 (Fragment).		
GN	MSp1.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
ON	NCBI_TaxID=5833;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=IFA9.26;		
PX	MEDLINE=20106724; PubMed=10643908;		
RA	Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,		
RA	Pluschke G.;		
RT	"Sequence diversity of the merozoite surface protein 1 of Plasmodium		
RT	falciparum in clinical isolates from the Kilombero District,		
RT	Tanzania.";		
BL	Acta Trop. 74:51-61(2000).		
DR	EMBL; AF061131; AAC69730.1; -.		
FT	NON_TER 347 347		
SQ	SEQUENCE 347 AA; 39603 MW; 5E155E74DB6DC0CB CRC64;		
Query Match 18.3%; Score 1544.5; DB 5; Length 347;			
Best Local Similarity 86.2%; Pred. No. 1.7e-41;			
Matches 307; Conservative 11; Mismatches 29; Indels 9; Gaps			
QY	1	MKIIFFLCSFLFFIINTQCVTTHESYQELVKKLEDALEDAVLTYSLFQEKRMVLNEGTSGT	60
Db	1	MKIIFFLCSFLFFIINTQCVTTHESYQELVKKLEDALEDAVLTYGLFHKKMILNEBEITT	60
QY	61	AVTTSTPGKGSVAGSGGSGSVAGSGSVAGSGSVAGSGSGNSRRTNFSDNSSDS	1200
Db	61	KGASAGSTGSGT--SGPSTGSGPSTGSPSSRSNTLPRSNFTSSAS-----PPADASDS	111
QY	121	DAKSADLKURVRNYLLTTIKELKYPQLFDLTNHLMTLCDNHGFVKYLDGYEINEELLYK	180
Db	112	DAKSADLKURVRNYLLTTIKELKYPQLFDLTNHLMTLCDNHGFVKYLDGYEINEELLYK	171
QY	181	LNIFYDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKLPLDNKDNVGKMD	2400
Db	172	LNIFYDLLRAKLDVNCANDYCOIPFNLIKIRANELDVLKLVFGYRKLPLDNKDNVGKMD	231
QY	241	YIKNNKTTIENINELIEESKTTIDKNNATKEEKKKLYOAOYDLSIYNKQLEEAHNLIS	3000
Db	232	YIKNNKTTIENINELIEESKTTIDKNNATKEEKKKLYOAOYDLSIYNKQLEEAHNLIS	2910
QY	301	VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKE	356
Db	292	VLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNTLLDKNKKIEEHEKEIKE	347
RESULT 54			

RT	Sequence	Query Match	Best Local Similarity	Matches	Indels	Gaps
RT	fclicparum in clinical isolates from the Kilombero District, Tanzania.;	18.3%; Score 1544; DB 5; Length 352;	86.0%; Pred. No. 1.9e-41;	19; Mismatches 23; Indels 8; Gaps		
RL	Acta Trop. 74:51-61(2000).					
RL	EMBL: AF061142; AAC69741.1; -.					
FT	NON_TER 352 352					
FT	SEQUENCE 352 AA; 39477 MW; LC62B4ED3025175B CRC64;					
QY	1 MKIIFFLCSELFIIINQCVTHESYQELVKKLEDAVLGTGSLFQKMKVINEGTSGT 60					
Db	1 MKIIFFLCSELFIIINQCVTHESYQELVKKLEDAVLGTGSLFQKMKVINEGTSGT 60					
QY	61 AVTTSTPGSGVSGGSGVSGGSGVSGGSGVSGGSGGSGNRRTNPSDSSDS 120					
Db	61 AVTTSTPGSGVSGGSGVSGGSGVSGGSGVSGGSGGSGNRRTNPSDSSDS 114					
QY	121 DAKSYADLKHVRNYLTIKELYPQLFDLTNNHMLTLCDDNIHGFYKYLIDGYEINELLYK 180					
Db	115 DAKSYADLKHVRNYLTIKELYPQLFDLTNNHMLTLCDDNIHGFYKYLIDGYEINELLYK 174					
QY	181 LNFYFDLLRAKLDVNCANDYQIPFNLIKIRANELDVLKLVFGYRKPDLNKNVCKMED 240					
Db	175 LNFYDILLRAKLDACANSYQIPFNLIKIRANELDVLKLVFGYRKPDLNKNVCKMED 234					
QY	241 YIKNKKTTIENIELIEESKTTIDKNKNATYKEBEKKLYQAOYDLSYNNKQLEAHNLIS 300					
Db	235 YIKNKKTTIENIELIEESKTTIDKNKNADNEGKKLYQAOYDLSYNNKQLEAHNLIS 294					
QY	301 VLEKRIDTLAKNENIKELLDKINEIK NPPANGSNTPNTLLD-KNKKIEBEHEKEIKE 356					
Db	295 VLEKRIDTLAKNENIKELLEDIDKIKTDAEKLITGTSKPNLPENKKKEVEGHEKEIKE 352					
RESULT 56						
QY	Q9TZV4 PRELIMINARY; PRT; 347 AA.					
AC	Q9TZV4;					
DT	01-MAY-2000 (TrEMBLrel. 13, Created)					
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)					
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)					
DE	Merozoite surface protein 1 (Fragment).					
GN	MSP1.					
OS	Plasmodium falciparum.					
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
OX	NCBI_TaxID=5833;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=IPAS.5;					
RX	MEDLINE=20106724; PubMed=10643908;					
RA	Jiang G., Daubenberger C., Huber W., Matile H., Tanner M.,					
RA	Pluschke G.;					
RT	"Sequence diversity of the merozoite surface protein 1 of Plasmodium					
RT	falciparum in clinical isolates from the Kilombero District,					
RL	Tanzania.;"					
RL	Acta Trop. 74:51-61(2000).					
DR	EMBL: AF061134; AAC69733.1; -.					
FT	NON_TER 347 347					
FT	SEQUENCE 347 AA; 39693 MW; 75DCA3237A9B7E33 CRC64;					
QY	Query Match	18.3%; Score 1542.5; DB 5; Length 347;				
Best Local Similarity	86.2%; Pred. No. 1.9e-41;					
Matches	307; Conservative 11; Mismatches 29; Indels 9; Gaps					
QY	1 MKIIFFLCSELFIIINQCVTHESYQELVKKLEDAVLGTGSLFQKMKVINEGTSGT 60					
Db	1 MKIIFFLCSELFIIINQCVTHESYQELVKKLEDAVLGTGSLFQKMKVINEEIT 60					
QY	61 AVTTSTPGSGVSGGSGVSGGSGVSGGSGVSGGSGGSGNRRTNPSDSSDS 120					
Db	61 KGASAOGTSGT--SGTSGSPGSGTSPSSRSRTLPRESNTSSGAS-----PPADASDS 111					


```
RN [2]
RP -SEQUENCE FROM N.A.
RX MEDLINE=95107347; PubMed=7808474;
RA Kerr P.J., Ranford-Cartwright L.C., Walliker D.;
RT "Proof of intragenic recombination in Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:241-248(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Ranford-Cartwright L.C.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X52963; CAA37136.1; -
FT NON_TER 1
FT NON_TER 363
SQ SEQUENCE 363 AA; 40336 MW; A86C08A93DE5AF9B CRC64;

Query Match      17.4%; Score 1467.5; DB 5; Length 363;
Best Local Similarity 81.6%; Pred. No. 4.4e-39;
Matches 298; Conservative 10; Mismatches 38; Indels 19; Gaps 4;

QY 21 THESYQELVKLEALEDAVLTCYSLFQKEKMWLNE---GTSCTAVTTSTPGSKGSVASGG 77
Db 1 THESYQELVKLEALEDAVLTCYSLFQKEKMWLNEEETTKGASAQSGASQSG--ASAQ 58

QY 78 SGGSVASGGSVASGGSVASGGSGN--SRRTN-----PSDNSSDSDAK 123
Db 59 SGASAQSGASQSGASQSGTSGPSGSGTSPSRSTLPRSTSSGASPPADASDSDAK 118

QY 124 SYADLKHRVRNYLLTIKELKYPQLFDLTNHMLTLCDNTHGFKYLDGYEINELLYKLN 183
Db 119 SYADLKHRVRNYLFTIKELKYPQLFDLTNHMLTLCDNTHGFKYLDGYEINELLYKLN 178

QY 184 YFDLLRAKLVNDYANDYCOIPENLIRANELDVLLKLVFGYRKPLDNKDNVGMEDYIK 243
Db 179 YFDLLRAKLVNDYANDYCOIPENLIRANELDVLLKLVFGYRKPLDNKDNVGMEDYIK 238

QY 244 KNKKTIENTNELIESKKTIDKNKNATKEEKKKLYQAOYDLSYINKOLEEAHNLISVLE 303
Db 239 KNKTTIANINELIEGSKKTIDQKNADNEEGKKKLYQAOYDLSYINKOLEEAHNLISVLE 298

QY 304 KRIDLKKNENIKELLDKINEIKNPPANGNTPTLLDKNKKIEEHEKEIKEIAKTIRF 363
Db 299 KRIDLKKNENIKELLDKINEIKNPPANGNTPTLLDKNKKIEEHEKEIKEIAKTIRF 358

QY 364 NIDSL 368
Db 359 NIDSL 363
```

Search completed: March 31, 2003, 07:25:26
Job time : 104 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 04:33:12 : Search time 25 Seconds
(without alignments)
2719.188 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKIIFFLCSEFLFFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	8424	100.0	1639	1	MSPI_PLAFW	P04933 plasmodium
2	8144.5	96.7	1630	1	MSPI_PLAFK	P04932 plasmodium
3	5136.5	61.0	1726	1	MSPI_PLAFPC	P04934 plasmodium
4	5127.5	60.9	1726	1	MSPI_PLAFPP	P50495 plasmodium
5	5069	60.2	1701	1	MSPI_PLAFM	P08569 plasmodium
6	5065	60.1	1701	1	MSPI_PLAFM	P13819 plasmodium
7	4854.5	57.6	1682	1	MSPI_PLAF3	P19598 plasmodium
8	2548.5	30.3	1772	1	MSPI_PLAFYO	P13828 plasmodium
9	1187	14.1	233	1	MSPI_PLAFD	P13827 plasmodium
10	1096.5	13.0	281	1	MSPI_PLAFN	P13820 plasmodium
11	501.5	6.0	1957	1	YD86_SCHPO	Q10411 schizosacch
12	486	5.8	1790	1	USO1_YEAST	P25386 saccharomyc
13	477	5.7	2869	1	RBP1_PLAVB	Q00798 plasmodium
14	474.5	5.6	1875	1	MLP1_YEAST	Q02455 saccharomyc
15	457.5	5.4	1251	1	RBP2_PLAVB	Q00799 plasmodium
16	443.5	5.3	2663	1	CENE_HUMAN	Q02224 homo sapien
17	440	5.2	1805	1	HMW2_MYCCE	P47460 mycoplasma
18	424	5.0	2230	1	GOG4_HUMAN	Q13439 homo sapien
19	423	5.0	1679	1	YIO9_YEAST	P40457 saccharomyc
20	421.5	5.0	1928	1	MSY1_YEAST	P08964 saccharomyc
21	405	4.8	2022	1	ANTI_ONCOC	P21249 onchocerca
22	404	4.8	2116	1	MSY2_DICDI	P08799 dictyosteli
23	399.5	4.7	1727	1	ALM1_SCHPO	Q9utk5 schizosacch
24	393	4.7	1312	1	RA50_YEAST	P12753 saccharomyc
25	386.5	4.6	2748	1	NUM1_YEAST	Q00402 saccharomyc
26	378	4.5	1818	1	HMW2_MYCPN	P75471 mycoplasma
27	372	4.4	1005	1	RA50_METJA	Q58718 methanococc
28	365	4.3	3911	1	AKA9_HUMAN	Q99996 h a-kinase
29	361	4.3	3660	1	DMD_CHICK	P11533 gallus gall
30	358.5	4.3	2349	1	TPR_HUMAN	P12270 homo sapien
31	355.5	4.2	3210	1	CENF_HUMAN	P49454 homo sapien
32	353	4.2	1169	1	SNC_METJA	Q59037 methanococc
33	352.5	4.2	1163	1	SBCCL_CLOAB	Q97fk1 clostridium

P02564	rattus norv	34	351.5	4.2	1935	1	MYH7_RAT
P12883	homo sapien	35	349.5	4.1	1935	1	MYH7_HUMAN
P34367	caenorhabdi	36	349.5	4.1	2198	1	YLJ2_CAEEL
P13539	mesocricetu	37	349	4.1	1939	1	MYH6_MESAU
P02565	gallus gall	38	348	4.1	1940	1	MYH3_CHICK
P32908	saccharomyc	39	347	4.1	1225	1	SMC1_YEAST
O67124	aquifex aeo	40	344.5	4.1	1978	1	RA50_AQUAE
P13540	mesocricetu	41	344.5	4.1	1934	1	MYH7_MESAU
P09975	marichantia	42	344	4.1	2136	1	YCF2_MARPO
P13533	homo sapien	43	343	4.1	1939	1	MYH6_HUMAN
Q02566	mus musculus	44	341.5	4.1	1938	1	MYH6_MOUSE
P10587	gallus gall	45	341.5	4.1	1978	1	MYH6_CHICK
O51578	borrelia bu	46	338	4.0	1169	1	EX5B_BORBU
P02563	rattus norv	47	338	4.0	1938	1	MYH6_RAT
Q03001	homo sapien	48	338	4.0	2704	1	BPA1_HUMAN
O15078	homo sapien	49	337.5	4.0	1539	1	Y373_HUMAN
P30622	homo sapien	50	336.5	4.0	1427	1	REST_HUMAN
P11532	homo sapien	51	335	4.0	3685	1	DMD_HUMAN
P15924	homo sapien	52	334.5	4.0	2871	1	DESP_HUMAN
P05661	drosophila	53	333.5	4.0	1962	1	MYSA_DROME
P79293	sus scrofa	54	332.5	3.9	1935	1	MYH7_PIG
Q15431	homo sapien	55	332	3.9	976	1	SCP1_HUMAN
Q97wn0	sulfolobus	56	331.5	3.9	864	1	RA50_SULSO
P24733	aequipekten	57	331.5	3.9	1938	1	MYS_ABOIR
P13538	gallus gall	58	331	3.9	1938	1	MYSS_CHICK
Q01550	xenopus lae	59	330.5	3.9	1744	1	TANA_XENLA
P54697	dictyosteli	60	330.5	3.9	2245	1	MYSJ_DICDI
Q03410	rattus norv	61	330	3.9	997	1	SCP1_RAT
Q9UKX3	homo sapien	62	326	3.9	1938	1	MYHD_HUMAN
Y373	BOVIN	63	324.5	3.9	1453	1	Y373_BOVIN
Q96YR5	sulfolobus	64	324	3.8	879	1	RA50_SULTO
P12882	homo sapien	65	323	3.8	1939	1	MYH1_HUMAN
P32380	saccharomyc	66	322.5	3.8	944	1	NUF1_YEAST
P28616	gallus gall	67	322.5	3.8	1102	1	MYSC_CHICK
Q28641	oryctolagus	68	322.5	3.8	1938	1	MYH4_RABIT
Q92351	schizosacch	69	321.5	3.8	1208	1	PCP1_SCHPO
P27625	plasmodium	70	320.5	3.8	2339	1	RPC1_PLAFA
P38989	saccharomyc	71	320	3.8	1170	1	SMC2_YEAST
P35580	homo sapien	72	320	3.8	1976	1	MYHA_HUMAN
Q99323	drosophila	73	320	3.8	2017	1	MYSN_DROME
Q62209	mus musculus	74	319.5	3.8	993	1	SCP1_MOUSE
Q33600	sulfolobus	75	319	3.8	886	1	RA50_SULAC
P58301	pyrococcus	76	318	3.8	882	1	RA50_PYRFU
P13535	homo sapien	77	317	3.8	1937	1	MYH8_HUMAN
O14157	schizosacch	78	317	3.8	2104	1	MYS3_SCHPO
Q94623	homo sapien	79	316.5	3.8	1939	1	MYH4_HUMAN
Q03661	saccharomyc	80	313	3.7	1658	1	YH67_YEAST
Q97592	canis famli	81	312.5	3.7	3680	1	DMD_CANFA
Q971t0	rattus norv	82	312	3.7	1976	1	MYHA_RAT
P12847	rattus norv	83	310	3.7	1940	1	MYH3_RAT
Q27991	bos taurus	84	309	3.7	1976	1	MYHA_BOVIN
Q04956	plasmodium	85	307.5	3.7	1956	1	ATX1_PLAFA
P02562	oryctolagus	86	306.5	3.6	1084	1	MYSS_RABIT
Q99104	mus musculus	87	306.5	3.6	1853	1	MYSA_MOUSE
Q60563	mesocricetu	88	305	3.6	845	1	SCP1_MESAU
Q24702	dictyocaulu	89	305	3.6	1557	1	DVAL_DICVI
Q49419	mycoplasma	90	304.5	3.6	756	1	Y328_MYCCE
P97929	mus musculus	91	303.5	3.6	3329	1	BRC2_MOUSE
P20929	homo sapien	92	303	3.6	6669	1	NEBU_HUMAN
P27895	saccharomyc	93	302.5	3.6	1038	1	CIN8_YEAST
P38198	saccharomyc	94	301.5	3.6	1513	1	STU1_YEAST
Q9cfz0	lactococcus	95	301	3.6	1046	1	SBCCL_LACLA
P11055	homo sapien	96	301	3.6	1940	1	MYH3_HUMAN
Q90988	gallus gall	97	300	3.6	1189	1	SCII_CHICK
Q90339	cyprinus ca	98	298.5	3.5	1935	1	MYSS_CYPCA
P16154	clostridium	99	298.5	3.5	2710	1	TOXA_CLODI
Q08372	plasmodium	100	298	3.5	3135	1	S230_PLAFO
Q42184	gallus gall	101	297.5	3.5	1433	1	REST_CHICK
P11531	mus musculus	102	296.5	3.5	3678	1	DMD_MOUSE
Q27171	paramecium	103	296.5	3.5	4540	1	DYHC_PARTE
P50532	xenopus lae	104	296	3.5	1290	1	XCPX_XENLA
P46939	homo sapien	105	295	3.5	3433	1	UTRO_HUMAN
Q11102	caenorhabdi	106	294	3.5	1130	1	YL17_CAEEL

180	256	3.0	743	1	ABRA_PLAFC	P22620	plasmodium
ALIGNMENTS							
RESULT 1							
MSPI_PLAFC	STANDARD;	PRT;	1639	AA.			
ID	MSPI_PLAFC						
AC	P04933;						
DT	13-AUG-1987	(Rel. 05, Created)					
DT	01-FEB-1996	(Rel. 33, Last sequence update)					
DT	01-FEB-1996	(Rel. 33, Last annotation update)					
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens) (PMMSA) (P195).						
GN	MSP-1.						
OS	Plasmodium falciparum (isolate Wellcome).						
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.						
OX	NCBI_TaxID=5848;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=86014355; PubMed=2995820;						
RA	Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V., Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T., Freeman R.R.;						
RA	"Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum merozoites.";						
RT	Nature 317:270-273(1985).						
RL	[2]						
RP	REVISIONS.						
RA	Holder A.A.;						
RL	Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.						
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).						
CC	-I- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).						
CC	-----						
DR	EMBL; X02919; CAA26676.1; --						
DR	PIR; A24594; A24594.						
DR	InterPro; IPR000561; EGF-like.						
DR	Pfam; PF00008; EGF; 1.						
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.						
KW	SIGNAL	1	19				
FT	CHAIN	20	1639	POTENTIAL. MEROZOITE SURFACE PROTEIN 1.			
FT	CARBOHYD	116	116	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	268	268	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	764	764	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	768	768	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	783	783	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	844	844	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	920	920	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	1058	1058	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	1165	1165	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	1174	1174	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	1445	1445	N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD	1526	1536	N-LINKED (GLCNAC. .) (POTENTIAL).			
SQ	SEQUENCE	1639	AA;	187618	MM;	2C255B6616C87F6E	CRC64;
Query Match 100.0%; Score 8424; DB 1; Length 1639;							
Best Local Similarity 100.0%; Pred. No. 2.9e-267;							
Matches 1639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							


```
QY 1 MKIIFFLCSFLFFIINTQCVTTHESQELVKVLEALEDAVLGTGYSLFQKEMVNLGTSQT 60
Db 1 MKIIFFLCSFLFFIINTQCVTTHESQELVKVLEALEDAVLGTGYSLFQKEMVNLGTSQT 60
QY 61 AVTTSTPGSKGVSAGSGGSGSVASGSGSVASGSGSVASGSGSGGSGNRRTNPDSNDS 120
Db 61 AVTTSTPGSKGVSAGSGGSGSVASGSGSVASGSGSVASGSGSGGSGNRRTNPDSNDS 120
QY 121 DAKSVADLKHVRNRYLLTIKELKYPOLFDLTNHLMTLCLDTHGFKYLDGYEIEINELLYK 180
Db 121 DAKSVADLKHVRNRYLLTIKELKYPOLFDLTNHLMTLCLDTHGFKYLDGYEIEINELLYK 180
QY 181 LNFYEDLLRAKLNDVANDYCOIPNKLIRANELDVLKLVFGYRKLPLDNTKDNVGRMED 240
Db 181 LNFYEDLLRAKLNDVANDYCOIPNKLIRANELDVLKLVFGYRKLPLDNTKDNVGRMED 240
QY 241 YIKKNNKTENINELIESKKTIDKNKATKEERKKLYQAOYDLSIYNKOLEEAHNLS 300
Db 241 YIKKNNKTENINELIESKKTIDKNKATKEERKKLYQAOYDLSIYNKOLEEAHNLS 300
QY 301 VLEKRIDTLKKNENIKELDKINEIKNPPANGSNTPNTLLDKNKKIEBEHEKEIKEIAKT 360
Db 301 VLEKRIDTLKKNENIKELDKINEIKNPPANGSNTPNTLLDKNKKIEBEHEKEIKEIAKT 360
QY 361 IKFNIDSFTDPLELEYLREKNNKIDISAKVETKESTEPNEYPNGVTYPLSYNDINNAL 420
Db 361 IKFNIDSFTDPLELEYLREKNNKIDISAKVETKESTEPNEYPNGVTYPLSYNDINNAL 420
QY 421 NELNSFGDLINFPDYTKPSKNIYTDNERKKFPIEIKIEKIEKKKIESDKKSYEDRSKS 480
Db 421 NELNSFGDLINFPDYTKPSKNIYTDNERKKFPIEIKIEKIEKKKIESDKKSYEDRSKS 480
QY 481 LNDITKEYEKLNEIYDSKFNNIDLTNFEKMMGRYKYVEKLTTHHTTFASEYNSKKNL 540
Db 481 LNDITKEYEKLNEIYDSKFNNIDLTNFEKMMGRYKYVEKLTTHHTTFASEYNSKKNL 540
QY 541 EKLTKALYMEDYSURNIWEKELYKNLISKIENEIETLVENIKKDEQLFKKITKD 600
Db 541 EKLTKALYMEDYSURNIWEKELYKNLISKIENEIETLVENIKKDEQLFKKITKD 600
QY 601 ENKPEKILEYSDIVKVOVKVLLMANKIDELKKTOLILKNVELKHNHVPNSYKOEKOE 660
Db 601 ENKPEKILEYSDIVKVOVKVLLMANKIDELKKTOLILKNVELKHNHVPNSYKOEKOE 660
QY 661 PYLLVLKKEIDKLVFMPKVESLINEBKNIKTGEGSDNSEPSTGEBITGOATTKPGQ 720
Db 661 PYLLVLKKEIDKLVFMPKVESLINEBKNIKTGEGSDNSEPSTGEBITGOATTKPGQ 720
QY 721 AGSALGDSVQAQAEQKQAQPPVPVPEAKAQVTPPAVPNNKTENVSKLDYLEKLYE 780
Db 721 AGSALGDSVQAQAEQKQAQPPVPVPEAKAQVTPPAVPNNKTENVSKLDYLEKLYE 780
QY 781 FLNTSYICHKYTLVSHSTWNEKILKOYKITKEESKLSGCDPLDLFLNQNINPVMSMF 840
Db 781 FLNTSYICHKYTLVSHSTWNEKILKOYKITKEESKLSGCDPLDLFLNQNINPVMSMF 840
QY 841 DSLNNSLSOLFMEIYEKEMVCNLYKLKDNKIDKNLLEAKVYSTSVKTLSSSMQPLSLT 900
Db 841 DSLNNSLSOLFMEIYEKEMVCNLYKLKDNKIDKNLLEAKVYSTSVKTLSSSMQPLSLT 900
QY 901 PQDRPEVSANDDTSHSTNLNLSKLFIENILSLGKNKIYQELIGKSSSENYEYKILKQSD 960
Db 901 PQDRPEVSANDDTSHSTNLNLSKLFIENILSLGKNKIYQELIGKSSSENYEYKILKQSD 960
QY 961 TFYNESFTNFVKSADDTINSUNDESKRKKLEEDINKLAKTTLQSLPDLYNKYKLERLFD 1020
Db 961 TFYNESFTNFVKSADDTINSUNDESKRKKLEEDINKLAKTTLQSLPDLYNKYKLERLFD 1020
QY 1021 KKKTVGKYMOKIKLTLKLEQLESKLSNLPKPHVLFQNSVFFNKKKEAEIAETENTLEN 1080
Db 1021 KKKTVGKYMOKIKLTLKLEQLESKLSNLPKPHVLFQNSVFFNKKKEAEIAETENTLEN 1080
QY 1081 TKILLKHYKGLVKKYNGESSPLKTLSESIQTEDNYASLENFKVLKLEGLKDNLNLEK 1140
```

```
Db 1081 TKILLKHYKGLVKKYNGESSPLKTLSESIQTEDNYASLENFKVLKLEGLKDNLNLEK 1140
QY 1141 KKLVSLSGLHLIAELAEVINKKNYTGNSPSENNTDVNNALLESYKKFLPEGTDVATVVS 1200
Db 1141 KKLVSLSGLHLIAELAEVINKKNYTGNSPSENNTDVNNALLESYKKFLPEGTDVATVVS 1200
QY 1201 ESGDFTLQSQPKPASTHVGAESNTITTSQNVDDVDDVIIVPIFGESEEDYDDLQGVV 1260
Db 1201 ESGDFTLQSQPKPASTHVGAESNTITTSQNVDDVDDVIIVPIFGESEEDYDDLQGVV 1260
QY 1261 TGEAVTPSVIDNLSKIENEYEVLYKLPLAGVYRSLKOLENNVMTFNVNVDILNSRN 1320
Db 1261 TGEAVTPSVIDNLSKIENEYEVLYKLPLAGVYRSLKOLENNVMTFNVNVDILNSRN 1320
QY 1321 KRENEKNVLESOLIPYKDLTSSNVVVKPYKFLNKEKRDKFLSSYNIKDSITDIDNFAN 1380
Db 1321 KRENEKNVLESOLIPYKDLTSSNVVVKPYKFLNKEKRDKFLSSYNIKDSITDIDNFAN 1380
QY 1381 DVLGYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFNNIETLYKTVNDKIDLFVHLE 1440
Db 1381 DVLGYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFNNIETLYKTVNDKIDLFVHLE 1440
QY 1441 AKVLNYYTEKSNVEVKIKELNYLKTIDQKLADFKNNNPFVGIADLSTDYNNHNNLTKFUS 1500
Db 1441 AKVLNYYTEKSNVEVKIKELNYLKTIDQKLADFKNNNPFVGIADLSTDYNNHNNLTKFUS 1500
QY 1501 TGMVFENLAKTVLSNLLDGNLQGLMNIHQHOCVKKOCPOKNSCFRHLDERECKCLLNK 1560
Db 1501 TGMVFENLAKTVLSNLLDGNLQGLMNIHQHOCVKKOCPOKNSCFRHLDERECKCLLNK 1560
QY 1561 QEGDKVCENPNPTCENNGGCDADAKCTEEDSGSNGKKTITCECTKPDSPSYPLFDGIFCSSS 1620
Db 1561 QEGDKVCENPNPTCENNGGCDADAKCTEEDSGSNGKKTITCECTKPDSPSYPLFDGIFCSSS 1620
QY 1621 NPLGISFLLILMLILYSFI 1639
Db 1621 NPLGISFLLILMLILYSFI 1639

RESULT 2
MSPI_PLAFK STANDARD: PRT: 1630 AA.
ID MSPI_PLAFK
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMWSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate KI / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID:5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```


the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; X03371; CAA37070.1; -	
DR	PIR; A25120; SAZQK1.	
DR	InterPro; IPR000561; EGF-like.	
DR	Pfam; PF00008; EGF; 1.	
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;	
KW	Transmembrane; GPI-anchor.	
FT	SIGNAL 1 19	POTENTIAL.
FT	CHAIN 20 1630	MEROZOITE SURFACE PROTEIN 1.
FT	DOMAIN 67 84	TRIPEPTIDE SG(TP) REPEAT.
FT	TRANSMEM 1614 1630	MEMBRANE ANCHOR.
FT	CARBOHYD 97 97	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 259 259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 755 755	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 759 759	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 774 774	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 835 835	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 911 911	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 955 955	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1049 1049	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1156 1156	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1165 1165	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1436 1436	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 1517 1517	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;	

Query Match 96.7%; Score 8144.5; DB 1; Length 1630;
Best Local Similarity 97.1%; Pred. No. 3.6e-258;
Matches 1591; Conservative 11; Mismatches 28; Indels 9; Gaps 2;

QY	1	MKIIFLCFLPIINTQCVTHESYQELVKKLEALDAVLTYGSLFQKEXKMWLNCTSGT	60
DB	1	MKIIFLCFLPIINTQCVTHESYQELVKKLEALDAVLTYGSLFQKEXKMWLNCTSGT	60
QY	61	AVTSTPGSGSVASGSGSVASGSGSVASGSGSVASGSGSRRTPNSDSSDS	120
DB	61	KGASGSGTSGT--STSGPSGSGTSPSRNTLPSRNTSSGAS-----PPADASD	111
QY	121	DAKSYADLKHVRNYLLTIKELYPQLFDLTNHLMTLCNIGHFKYLDIGYEBINELLYK	180
DB	112	DAKSYADLKHVRNYLLTIKELYPQLFDLTNHLMTLCNIGHFKYLDIGYEBINELLYK	171
QY	181	LNFPDILLRAKLVNDVANDYCOIPFNLIKTRANELDVLLKLVFGYRKPDLNDKVNKMD	240
DB	172	LNFPDILLRAKLVNDVANDYCOIPFNLIKTRANELDVLLKLVFGYRKPDLNDKVNKMD	231
QY	241	YIKKKNKTTENINELLEESKTTIDKKNATKEBEKKLYQAOYDLSIYNKQLEAHNLIS	300
DB	232	YIKKKNKTTENINELLEESKTTIDKKNATKEBEKKLYQAOYDLSIYNKQLEAHNLIS	291
QY	301	VLEKRIDTLKKENIKELLDKINEIKNPPANSNGTNPNTLLDNKKIEEHEKEIKEIAKT	360
DB	292	VLEKRIDTLKKENIKELLDKINEIKNPPANSNGTNPNTLLDNKKIEEHEKEIKEIAKT	351
QY	361	IKFNIDSLTDPLEYLYIREKNKNIDISAKVETKESTPEPNVPGVTPLSYNDINNAL	420
DB	352	IKFNIDSLTDPLEYLYIREKNKNIDISAKVETKESTPEPNVPGVTPLSYNDINNAL	411
QY	421	NELNSFGDLINPDYTKPSKNITYDNERKKFNEIKETKIEKKIESDKKSYEDRSKS	480
DB	412	NELNSFGDLINPDYTKPSKNITYDNERKKFNEIKETKIEKKIESDKKSYEDRSKS	471
QY	481	LNDDITKEYEKLLEIYDSKFNENNIDITNPEKMMGRYSKYKVEKLTHHNTFASYSKHN	540
DB	472	LNDDITKEYEKLLEIYDSKFNENNIDITNPEKMMGRYSKYKVEKLTHHNTFASYSKHN	531
QY	541	EKLTAKLYMEDYSLRNIVVEKELKYKKNLISKIENIEITLVENIKKDEQLEFEKKITKD	600

DB	532	EKLTAKLYMEDYSLRNIVVEKELKYKKNLISKIENIEITLVENIKKDEQLEFEKKITKD	591
QY	601	ENKPDKEILEVSDIVAVQVKVLLMKNKIDELKKTQILTKNVELKHNHIVPNSYKQENKQE	660
DB	592	ENKPDKEILEVSDIVAVQVKVLLMKNKIDELKKTQILTKNVELKHNHIVPNSYKQENKQE	651
QY	661	PYYLIVLKKKIDKLVFMPKVESLINEEKKNIKTEQSDNSSEPTGEITGQATTPGQQ	720
DB	652	PYYLIVLKKKIDKLVFMPKVESLINEEKKNIKTEQSDNSSEPTGEITGQATTPGQQ	711
QY	721	AGSALEGDSVQAQAQEQKQAPPPVPVPEAKAQVTPTPAPVNNKTNENSKLDYLEKLYE	780
DB	712	AGSALEGDSVQAQAQEQKQAPPPVPVPEAKAQVTPTPAPVNNKTNENSKLDYLEKLYE	771
QY	781	FLNTSYICHKYILVSHSTWNEKILKOYKITKEESKLSSCDPLDLLFNIONNIPVMYSMF	840
DB	772	FLNTSYICHKYILVSHSTWNEKILKOYKITKEESKLSSCDPLDLLFNIONNIPVMYSMF	831
QY	841	DSLNNLSQLFMEIYEKEMVCNLYKLDNDKIKNLLLEAKKVSTSVKTLSSSSMQPLSLT	900
DB	832	DSLNNLSQLFMEIYEKEMVCNLYKLDNDKIKNLLLEAKKVSTSVKTLSSSSMQPLSLT	891
QY	901	PODKPEVSANDTSHSTNLSNLSKLFENILSGKNKIYQELIGQKSSSENFYEKILKDS	960
DB	892	PODKPEVSANDTSHSTNLSNLSKLFENILSGKNKIYQELIGQKSSSENFYEKILKDS	951
QY	961	TFYNESFTNFVSKADDDINSNDKSKKLEEDINKLKTQLSFDLYNKYKLLERLFD	1020
DB	952	TFYNESFTNFVSKADDDINSNDKSKKLEEDINKLKTQLSFDLYNKYKLLERLFD	1011
QY	1021	KKTVCKYKMQIKKTLTLKEQLESKLSNPNKHVLQNFVSFFENKKKEAIEAETENTEN	1080
DB	1012	KKTVCKYKMQIKKTLTLKEQLESKLSNPNKHVLQNFVSFFENKKKEAIEAETENTEN	1071
QY	1081	TKILLKHVGLVYKNGESSPLKTLSEESTQTEDNTASLENFKVLSKLEKLDNLEK	1140
DB	1072	TKILLKHVGLVYKNGESSPLKTLSEESTQTEDNTASLENFKVLSKLEKLDNLEK	1131
QY	1141	KKLSYSSGLHHLIAELKEVKNKNTGNSPSNNTDVNALESYKFFPEGTDVATVS	1200
DB	1132	KKLSYSSGLHHLIAELKEVKNKNTGNSPSNNTDVNALESYKFFPEGTDVATVS	1191
QY	1201	EGSDTLEQSQPKKPASTHVGAESENTITTSQNVDDDEVDDVIIIPFGESEEDYDLGQVV	1260
DB	1192	EGSDTLEQSQPKKPASTHVGAESENTITTSQNVDDDEVDDVIIIPFGESEEDYDLGQVV	1251
QY	1261	TGEAVTPSIDNTLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVNKKDILNSRFN	1320
DB	1252	TGEAVTPSIDNTLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVNKKDILNSRFN	1311
QY	1321	KRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKROKFLSSYNYIKDSIDTDINFA	1380
DB	1312	KRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLNKEKROKFLSSYNYIKDSIDTDINFA	1371
QY	1381	DVLGYKILSEKYSKLDSDIKKYINDKQENKYLPLFNNIETLYKTVDNDKIDLFVHLE	1440
DB	1372	DVLGYKILSEKYSKLDSDIKKYINDKQENKYLPLFNNIETLYKTVDNDKIDLFVHLE	1431
QY	1441	AKVLNVTYKSNVVEKIKELNYLKTQDLADPKKNNFVGADLSTDYNNHLLTKFLS	1500
DB	1432	AKVLNVTYKSNVVEKIKELNYLKTQDLADPKKNNFVGADLSTDYNNHLLTKFLS	1491
QY	1501	TGMVFENLAKTVLSNLDGNLQGLMNIHQCVKQKQPCQNSGCFRHLDERECKCLLNYK	1560
DB	1492	TGMVFENLAKTVLSNLDGNLQGLMNIHQCVKQKQPCQNSGCFRHLDERECKCLLNYK	1551
QY	1561	QEGDKCVENPNTCNENNGGCDADAKCTBEDSGSNKKITCECTKPDSPFLPDGIFCSSS	1620
DB	1552	QEGDKCVENPNTCNENNGGCDADAKCTBEDSGSNKKITCECTKPDSPFLPDGIFCSSS	1611
QY	1621	NFLGISFLLILMLILYSFI 1639	
DB	1612	NFLGISFLLILMLILYSFI 1630	

Qy	561	EKELYKYKNLISKIENIEITLVENIKKKDEQLPEKK---ITKDENKPDEKILEVSDIVKV	617
Db	594	EKFNFHYTTLKTGLEADIKLTBEIKSSSENKLEKNFKGLTHSANAS----LEVYDITVKL	649
Qy	618	QVQVLNMNKIDELKTLTKLVNKLKHNIHPVNSYKOKENKOEPYLLIVLAKKEDKLKVP	677
Db	650	QVQVLIIKKIEDLRKIELFLKNAQKDSIHVPNIYKQNPKPEPYLLIVLAKKEVDLUKEF	709
Qy	678	MPKVESLINEEKNKIETEQOSONSEPSTEGETGATTKPGQAQSALSGSVOAQAEQ	737
Db	710	IPVKDMLKKEQAVLSS-----ITQPLVAASETTEDGGHSHTLTSQSGETEVTEET	760
Qy	738	KQAOPPV-----PVPVPEAKAQPVTTPAPVNNKNTENVSKLDYLEKLYEFELNTSYICH	789
Db	761	EETEETVGHTTVTITLPPEKVKVWENSIEHKSNDNSOALTKTYYVLKKLDFELTKSYICH	820
Qy	790	KYILVSHSTWNEKILKOYKITKEERSKSSCDPLDLLENIONNIIPVMYSMDFSNNLSLSQ	849
Db	821	KYILVSNSMDQLLEYNLPTPEEENELKSCDPOLLFFLNQNNIPAMISLYDSMNNDLQH	880
Qy	850	LPMEIYEKEMCVNLKYLKDNKDIKNLLEBAKV-----STSVAKTLSSSS-----	893
Db	881	LFPELYOKEMYIYHLKLEENHIKKLLEBOKOITCTSSSPGNITVMTAQSAHNSQN	940
Qy	894	--MQPLSLTPQDRPEVSAND---DTSSH---TNLNSLKLFEINILSGKNKNITYQEL-I	943
Db	941	QQSNASSNTNQNGVAVSSGPVAVEESHDPULVLSISNDLKGIVLSLLNGLNKTVPNPULTI	1000
Qy	944	GOKSSENFYEKILKDSOTYNESFTNVFYKSKADDINSLNDESRKKLEEDINKLKTLOL	1003
Db	1001	STTEMKFIYENILKNDDTVFNDDIKQFVKSNSKVITGTU-ETQRNALNDELTKLKUDTLQOL	1059
Qy	1004	SFDLYNYKILKLERLDFOKKTKTYGVYKMQIKTLTLKEQLESKLSLNNPKHVLFQNSVFV	1063
Db	1060	SFDLYNKYKLDRLFNKKKELGDQMJKLTLTLKEQLESKLSLNNPHVNLFQNSVFV	1119
Qy	1064	NKKKEABIATENTLNTKILKHKYKGLVYKYNGESSPLKTLSEESTQTEDNYASFENFK	1123
Db	1120	NKKKEABIAETENTLNTKILKHKYKGLVYKYNGESSPLKTLSEYSTQEDNYANLEKFR	1179
Qy	1124	VLSKLEGKLDNLNLEKKLSYSSGLHHLIAELKEVIKNKNYTGNSPSENNTOYNALE	1183
Db	1180	VLSKDGKLDNLNHLGKKKLSFLSSGLHQILTELKEVINKNKNYTGNSPENKNKVNEALK	1239
Qy	1184	SYRKFLPEGTDVATVYSE-----SGSDTLQSQPKPPASTHVGAESNTITT	1229
Db	1240	SYENFLPE-AKVITVTVTPQDPVTPSPLSVRVSGSGSTKEETQIPTSGSLLELOQVVQ	1298
Qy	1230	SONYDEVDDVIIVPIFGSEEDYDDLGOVVTGEAVTPSVINDTLISKIENEYLYLKLPL	1289
Db	1299	LQNYDEDDSLVPLIFGESEDNDEYLDQVVTGEAISVT-MDNILSGFENEYDYLYLKLPL	1357
Qy	1290	AGYVRSCLKOLENNVTFPNVKDILNSRPNKRENFKNVLESDDLIPYKDLTSSNYVVVKDP	1349
Db	1358	AGYVRSCLKOIEKNIFTNLNLDNLNSRLKRKYFLDVLEDSLMOFKHITSNEYIEDS	1417
Qy	1350	YKELNKEKRDCKFLSSNYIKDSITDIDINFANDVLGYKYLSEKVKSDLDSITKKYI----	1404
Db	1418	FKLLNSEQKNTLKSXYIKESVENDIKFAQBGISYIEBKVIARYKDDULESKVKIKEEKE	1477
Qy	1405	-----NDKOGENEKYLPLFNNTITETYKTVNDKIDLFIHLEAKVLNITYE	1449
Db	1478	KFPSSPPTTPPSPAKDEQKESKFLPFLTNETILYNNLVNKIDDYLINLAKINCDCNVE	1537
Qy	1450	KSNVEYKIKELANYKTIQDKLADFAPKNNNFVGIADLSTDYWHNNLLTKFISTGMGFENLA	1509
Db	1538	KDEARHKITKLSDLKAIDDKIDLFFNHNDFOAIKKLIINDDTFKMDGLKLISTGLV-QNPF	1596
Qy	1510	KTVLSNLLDGNLGMLNISOHOCVKQCPONSGCFRHLDERECECKCLINTYKQEGDKVCEN	1569
Db	1597	NTIISKLIEGKFQDMNLNISQHCVKQCPENSGCFRHLDERECECKCLINTYKQEGDKVCEN	1656

QY	1570	PNPTCNENNGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFGLGISFLL	1629
DB	1657	PNPTCNENNGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNFGLISFLL	1716
QY	1630	IILMLIYSFI	1639
DB	1717	IILMLIYSFI	1726
 RESULT 5 MSPI_PLAFM			
ID	MSPI_PLAFM	STANDARD;	PRT; 1701 AA.
AC	P08569;		
DT	01-AUG-1988 (Rel. 08, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	Merozoite surface protein 1 precursor (Merozoite surface antigens)		
DE	(PMSEA) (P190).		
GN	MSP-I.		
OS	Plasmodium falciparum (isolate mad20 / Papua New Guinea).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=70153;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88011243; PubMed=3079521;		
RA	Tanabe K., Mackay M., Goman M., Scaife J.G.;		
RT	"Allelic dimorphism in a surface antigen gene of the malaria parasite plasmodium falciparum";		
RL	J. Mol. Biol. 195:273-287(1987).		
CC	[2]		
RN	REVISIONS TO 1403; 1569 AND 1629.		
RA	Tanabe K.;		
RL	Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.		
RN	[3]		
RP	SEQUENCE OF 1-115 FROM N.A.		
RX	MEDLINE=86136024; PubMed=3004972;		
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,		
RA	Stunnenberg H., Bulard H.;		
RT	"Polymorphism of the precursor for the major surface antigens of plasmodium falciparum merozoites: studies at the genetic level.";		
RL	EMBO J. 4:3823-3829(1985).		
CC	-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (potential).		
CC	-!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@lsb-sib.ch).		
CC	EMBL; X05624; CAA29112.1; --		
DR	PIR; A26868; A26868.		
DR	PIR; B25120; B25120.		
DR	IncerPro; IPR000561; EGF-like.		
DR	Fam; PF00008; EGF; 1.		
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.		
FT	SIGNAL	1	19
FT	CHAIN	20	1701
FT	CARBOHYD	110	110
FT	CARBOHYD	239	239
FT	CARBOHYD	470	470
FT	CARBOHYD	536	536
FT	CARBOHYD	607	607
FT	CARBOHYD	802	802
FT	CARBOHYD	899	899
FT	CARBOHYD	919	919
FT	CARBOHYD	965	965
FT	CARBOHYD	965	965

RT of Plasmodium falciparum. *;
 RL Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -L- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (-, Potential).
 CC -L- PTH: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M19143; AA29653.1; -
 DR PIR: A54498; A54498.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF00008; EGF; 1.
 KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;

Query Match 60.1%; Score 5065; DB 1; Length 1701;

Best Local Similarity 60.1%; Pred. No. 5,2e-158;

Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;

QY 1 MKIIIFLCFLFFIINTQCWTHESYQELVKLEDAVLITGYSLFQKEKVMVNECTSGT 60
 DB 1 MKIIIFLCFLFFIINTQCWTHESYQELVKLEDAVLITGYSLFQKEKVMVNECTSGT 60
 QY 61 AVTTSTPGSGSGVSGGSGVSGGSGVSGGSGVSGGSGGSGNSRRTNPSONSSDS 120
 DB 61 AVTTSTPGSGSGVSGGSGVSGGSGVSGGSGGSGGSGGSGNSRRTNPSONSSDS 114
 QY 121 DAKSYADLKHRYNRLTITKELKYPOLFTLNHMLTCLDNIHGFYKYLIDGYEIEINELLYK 180
 DB 115 NTKTYADLKHRYNRYLFTITKELKYPOLFTLNHMLTCLDNIHGFYKYLIDGYEIEINELLYK 174
 QY 181 LNFYDOLLRAKLVANDCYOIPNKLIRANELDVLKLVFGYRKPLDNIRKDNVGMED 240
 DB 175 LNFYDOLLRAKLVANDCANSYCIPIPNKLIRANELDVLKLVFGYRKPLDNIRKDNVGMED 234
 QY 241 YIKKNNKTITENELIEESKKTIDKNKATKEEKKKLYQAQYDLSIYNKQLEEAHNLIS 300
 DB 235 YIKKNNKTITANLEIEGSKKTIDQKNADNEEGKKLYQAQYNLFYIYNKQLEEAHNLIS 294
 QY 301 VLEKIDTLTKNNETIKELADKINEIK- ---NPPANGNTPNTLLD-KNKKTEEHEKEIK 355
 DB 295 VLEKIDTLTKNNETIKELADKINEIK- ---TTGSKPNPLPENKKEVEGHEEKEIK 351
 QY 356 EIAKTIKFNIDSLTDPLEYLEYLRKKNIDISAKVE--TKESTEPN-EYPNGVTYPLS 412
 DB 352 EIAKTIKFNIDSLTDPLEYLEYLRKKNKVDVTPKSQDPTKSVQIPKVPYIPNGIVPLP 411
 QY 413 YNDINNAL- ---NELNSFGDLINPFDYTKPSKNIVTDN-ERKKFNEIKEIKTEKKKI- 467
 DB 413 YNDINNAL- ---NELNSFGDLINPFDYTKPSKNIVTDN-ERKKFNEIKEIKTEKKKI- 467

DB 412 LTDIHNSLAADNDKNSYGDLMNP-DTKEKINEKIIITDNKERKIFINNIKKQIDLEEKIN 470
 QY 468 ---ESDKKSYEDRSKSLNDITKEYEKLLEIYDSKFNENNIDLTNFEKMMKGRYSYVEKL 524
 DB 471 HTKEQNKKLLDEYKS- ---KKDYELLEKFYEMKFNNNFDDKVDVDFISARTIYNVQK 526
 QY 525 THHTFASYSKHNLEKLTAKLYMEDYSLRNIVVYVEKELKYKYLKLSIENIETLVEN 584
 DB 527 RYNNKFSSNNNSVYVQKLLKALSYLEYSLRKGISEKDFNHYITLTKGLEADIKKLTTEE 586
 QY 585 IKKDEBQLFEKK- ---ITKDENKPDKEILEVSDIVKVOVKVLMNKLDELKTKTQLILKNV 641
 DB 587 IKSENKILEKFNKGLUHSANAS- ---LEVSDIVKVOVKVLLIKKIDELKTELFLKNA 642
 QY 642 ELKHNIHVNSYKQENKQBPYLYLVKLEIDKLKVPKVESLINEKKNIKTEGQSDNS 701
 DB 643 QLKDSIHVNIYKPNKPEPYLYLVKLEIDKLKVPKVESLINEKKNIKTEGQSDNS 695
 QY 702 EPSTEIGTQATTKPGQAGSALGDSVQAQAEQKQA- ---QPPVPVPVPEAKAQ 754
 DB 696 --ITQPLVAASETDEGGHSTHTL-SQSGETEVEETEVEETVGHVTTITLPPKEES 752
 QY 755 VTPPPAPVNNKTEN- ---VSKLDYLEKLYEFLNTSYICHKYILVSHSTNNEKILQ 806
 DB 753 APKEVKVVENSIHKSNDNSQALTKTVYLLKLDLEFTKSYICHKYILVSNSSMDQKLEV 812
 QY 807 YKITTKEESKLSKSCDPLDLLENQNNIPVMSYFOSLNNLSOLFMEIYEKEMVCNLYKL 866
 DB 813 YNLTPEEENELKSCDPLDLLENQNNIPVMSYFOSLNNLSOLFMEIYEKEMVCNLYKL 872
 QY 867 KNDKIKNLEEAQKY- ---STSVKTLSSSS- ---MQPLSLTPQDRPEVS 908
 DB 873 KEENHKKLLEEQKQITGTSTSSPGNTVNTAQSATHSNQSNQSNASSTNTQNGVAVS 932
 QY 909 AND- ---DTSHS- ---TNLNNLSKLPENILSLGNKNKIYQEL-IGOKSSSENFYEKILKDS 960
 DB 933 SGPAVVEESHDLTVLISLNDLKGIVSLNLGNKTKVPNPLTISTTEMEKFFYENILKND 992
 QY 961 TPNESFTNFVKSADDINSLNDESKRKLLEEDINKKTLQSLDFLYNKKYKLERLFD 1020
 DB 993 TYFNDDIKOFVKSNSKVIKLT-ETOKNALNDEIKKLTQSLDFLYNKKYKLERLFD 1051
 QY 1021 KKKTVGKQMKIKKLTLLKEQLESKLNPNKHVLFQNFVFPFNKKKEAEIAETENTLEN 1080
 DB 1052 KKKELGQDKMQIKKLTLLKEQLESKLNPNKHVLFQNFVFPFNKKKEAEIAETENTLEN 1111
 QY 1081 TKILLKHGKLYKYNNGESSPLKLTSEESIOTEDNVTASLENFVLSKLECKLADNLEK 1140
 DB 1112 TKILLKHGKLYKYNNGESSPLKLTSEESIOTEDNVTASLENFVLSKLECKLADNLEK 1171
 QY 1141 KKLVSLSGLHLLIAELKEVKNKNTGNSPSENNTDVNNALESYKFFLPEGTDVATVVS 1200
 DB 1172 KKLVSLSGLHLLIAELKEVKNKNTGNSPSENNTDVNNALESYKFFLPEGTDVATVVS 1230
 QY 1201 E- ---SGSDLEQSQPKPASTHVAESNTITTSQNVDEVDDVIVPIF 1246
 DB 1231 PPQPDVTPSPSVRSVSGSGSTKEETQIPTSGSLLTQVVOLOQYDEEDSLVWLPIF 1290
 QY 1247 GSEEDYDDLQGVVTCGAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLLKQLENNVMT 1306
 DB 1291 GSEEDYDDLQGVVTCGAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLLKQLENNVMT 1349
 QY 1307 FNVNVDIENSRFNKREKNVLESOLIPYKDLTSSNVYKDPYKFLNKKRDKRKFSSYN 1366
 DB 1350 FNVNVDIENSRFNKREKNVLESOLIPYKDLTSSNVYKDPYKFLNKKRDKRKFSSYN 1409
 QY 1367 YIKDSTDTDINFANDVLYGKYLSEKYSKSDLSIKKYI- ---ND 1406
 DB 1410 YIKESVENDIKFAQEGISYVEKVLAKYKDDLESIKKVIKEKEKFPSSPPTPPSPAKTD 1469
 QY 1407 KOGENEKYLPLFNLIETLYKTNDKIDLFVHLEAKVLYNTEKSNVEKIKELNLYKTI 1466
 DB 1470 EOKESKFLPFTNIETLYNLYNKKIDDLINLAKINDCNVEKDEAHVKITKLSOLKAI 1529


```
QY 335 N-----TPN-----TLIDKN 344
: : : : :
Db 303 SSASGTSAGAGTVEQANTVASVTVPVSVGCEASTNPQTAQVQVPVTLILEEQ 362
: : : : :
QY 345 KKEEHEKEKEIAKTIKFNIDSLFTDPLBELEYLREKNK---NIDISA--KVETKESTE 399
: : : : :
Db 363 KKIAGLQAQKEIAKTIKFNIDSLFTDPLBELEYLREKNK---NIDISA--KVETKESTE 422
: : : : :
QY 400 P--NEYPNGVYPLSYNDI-----NNAELNSFGDLINFDYTKPESKNIYTDNERKKF 452
: : : : :
Db 423 PLTRIPNGISYPLPENDVYVYKIANNAE---TTYGDTHP-DNTPLTGDATNEQAKDL 479
: : : : :
QY 453 INETKIKIKIEKKIESDKSYDRSKSLNDITKEVEKLLNEIYDSKFNNDIDLTNPEKM 512
: : : : :
Db 480 IKAIKKIAEKKLEFLTKYNDKLTNTEFNOQKTPFEAAKEFESEFRKNLTSEIFEK 539
: : : : :
QY 513 MGRKYSYKVEKLTHHNTFASYENSKHNLEKLTKALKYMEDYSURNIVVEKELKIYKNLIS 572
: : : : :
Db 540 KTRDEYMTKK--TELNT--CEYGNTEKELINKLNQNLQYSLRKDIISNEIEYFSNKK 597
: : : : :
QY 573 KIENEIETLVENIKKDEQLFEKIKTKDENKPDKEKILEVSDIVKQVOKVLLMKNIDELK 632
: : : : :
Db 598 ELQTNINRLAEAVOAKONVLA-----SKOVPLSTLVLELOQKSLLTQKIQEQLN 646
: : : : :
QY 633 KTQILIKNVELKNIHVPNSYKOENKOEYLLIVLKEIDKLVFMPKVESLINEEKNI 692
: : : : :
Db 647 KTEVSLNKAQKDKLYPKTYGNEGKPEPYLLAVKKEVDRLAQFIPKIESMIKAKERM 706
: : : : :
QY 693 K-----TEGQSDNSEPSTEGEITGOATTKPQOQAGSA-----LEG 727
: : : : :
Db 707 EOGPAITGESEVPSPSAESSTDRSTQSSSTSSSSSTPAAAESSTATPEAPAPAE 766
: : : : :
QY 728 DSVQAOAQEQKQAPVPPVPEAKAQVPPAPVNNKTENVSKLDYLEKLYEFLNTSYI 787
: : : : :
Db 767 ASRSTEASETTTPPTQETQPSOASSTTPAKPV-----MTKLYLEKQLKQFLVFSYS 820
: : : : :
QY 788 CHRYLVSHSTMEKILQYKIKTEES---KLSSCDPLDLFLNIQNNIPVYVSNFDSLNN 845
: : : : :
Db 821 CHYVLLQNSTINKDALSRYALTEEDKIRTLARCSLDELVAIQNNMPTMYSIYESIVD 880
: : : : :
QY 846 SLQOLFMEIYKEMVNCVLYKLD--NKKIKNLEEA-----879
: : : : :
Db 881 GLQNIYELYKEMVHYIKLKDNPENSIKSLLYKAGVIEPEVPAAPTPTVPAATQEQOQ 940
: : : : :
QY 880 --KRVTSVTKLSSSMOPLSLTPQDKPEVSNDDTSHS-----916
: : : : :
Db 941 ATPDVQSDAPADSVQSQPETPTVTSTPTTEASSAPGEGTSGEAGASGTGATA 1000
: : : : :
QY 917 -----TNLNNSLKLFENILS-----LGKKNKIYQELI 943
: : : : :
Db 1001.SNAATPAGTSAGSAGSASNASTTSVDTPPAAAAAASPSTSTPAPAPPAANSQSGNPDGIR 1060
: : : : :
QY 944 GQKSSE-----NFEKILKSDTPTYNESFNFVKSKADDINSLNDESKRKLKEE 992
: : : : :
Db 1061 SRAESEDMPADDFELDNLYKSVLQIQDG--NNTFEINFIKSKKELIKALTPK--KVNLQYL 1118
: : : : :
QY 993 DINKLKTLLSFDLYNKYKLERLFDKKTGVKMOIKLTLKLEQLESKLSLNNP 1052
: : : : :
Db 1119 EIAHLKELSHYDRTYTKLRLYKHEQIQLNRQIRDLUSILKARLKKRQTLNGV 1178
: : : : :
QY 1053 KHYLVNFSVFNKKEAETAEENTLENTKILKHYKGLVYNGBSSPLKLTSEESIQT 1112
: : : : :
Db 1179 FYILNGVYVFNKREAREKQYVDNALKNITDMLLYKARTKYTSEAVPLKTLKSLDR 1238
: : : : :
QY 1113 EDNYASLENFKVLSKLEGLKONLNLEKKLSYLSGLHLHIAELKEVKNKNTGNSPS 1172
: : : : :
Db 1239 ESNLYKIEKFRAYSRLERLKLNKINLGKIRISYVSGLHVFEFEKLIKDKDYTKKNP 1298
: : : : :
QY 1173 ENNTDVNNALESYKFLPEGTQVATVVSSEGSTLTLEQSQPKPASTHVGAESNTI-----1227
: : : : :
Db 1299 DNAPEVTNAEQKELLPKGVTVST--PAVAVTTTLAADAATPEGAVPGVAVPGVAVP 1357
: : : : :
```

```
QY 1228 -----TTSQNVDDDEVDDVIIPIFGESEEDYDLDGQVVTGEAVTPSPVIDNLSK 1276
: : : : :
Db 1358 GAVPGVPGSGTDRVAGSSVDD-----NED--DDIYQIASQSEDAPEKD--ILSE 1405
: : : : :
QY 1277 IENEYEVLYLKLPLAGYVRSLSKQLENNVMTFNVNVDILNSRFNKNFKNVLESOLIPY 1336
: : : : :
Db 1406 FTRESYVYVTKRLGSTYKSLKHLMLREFSTIKEDMTNGLNKNKSKRNDLFLEVLSHELD 1465
: : : : :
QY 1337 KDLTSSYVVKDQPKFLNKKRDKFELSSYNYIKDSITDINFANDVLGYVYKILSEKYSK 1396
: : : : :
Db 1466 KDLSTNKYVIRNPYQLLNDKDKQVNLKAYATKGINEDIETTTDGIKFNKKNVLYNTQ 1525
: : : : :
QY 1397 LDSIKYI-----ND--KOGENEKYLPLNNIETLYKTVNDKIDLFVHLEAKVNLNTY 1448
: : : : :
Db 1526 LAAVEQOIAETAEETNDTNKEKKKYPILEDLKLGLYETVIGQAEYSEELQNRLDNYKN 1585
: : : : :
QY 1449 EKSNEVVKIKELYLTKIQDKLADF-----KKNNFVGIADLSTDYNNHNLTLTFLSTGMV 1504
: : : : :
Db 1586 EKAEEFILTAKNLEKYIQIDKLEDFEVEHAENKHIASIA-----LNNLNKSGLV 1634
: : : : :
QY 1505 FENLAKTVLSNLDGNLQGM--LNIS--OHOCV--KOCOPNSGCFRHLDEREECKLLNYK 1560
: : : : :
Db 1635 GEGESKILAKML--NMDGMDLLGVDPKHVCVDTROIKNAGCFRDNNGTEWRCLLYGK 1692
: : : : :
QY 1561 Q--EGDKRCVENPNTNENNGCGDADAKCTEEDSGSKKITCTECPDPSYPLDFGIFCSS 1619
: : : : :
Db 1693 KGBENTCVENNNPTCDINNGCGDPTASQNAESTENSKKICTCKEPTPNAYYEGVFCSS 1752
: : : : :
QY 1620 SNFLGHSFLILMLIYS 1637
: : : : :
Db 1753 SSPMGLSILLIITLIVFN 1770
: : : : :
RESULT 9
MSPL_PLAFD STANDARD; PRT; 233 AA.
AC P13827;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
DE (PMSA) (Glycoprotein 185) (GP185) (Fragment).
GN MSP-1.
OS Plasmodium falciparum (isolate CDC / Honduras).
OC Bukaryaota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87106855; PubMed=3542719;
RA Howard R.F., Ardeshir F., Reese R.T.;
RT "Conservation and antigenicity of N-terminal sequences of GP185 from
different Plasmodium falciparum isolates.";
RL Gene 46:197-205(1986).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(potential).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```


5	SEQUENCE	281 AA;	31226 MH;	6150B76E0F21B299	CRC64;	
	Query Match	13.0%	Score 1096.5;	DB 1;	Length 281;	
	Best Local Similarity	77.7%;	Pred. No. 8e-30;			
	Matches	227;	Conservative	8;	Mismatches 32;	Indels 25; Gaps 5;
QY	1	MKIIFFLCSEFLFIINTQCVTHESYQELVKKLEALEDAVLGTGSLFOKEKMYLNEGTS	60			
DB	1	MKIIFFLCSEFLFIINTQCVTHESYQELVKKLEALEDAVLGTGSLFOKEKMYLNE	56			
QY	61	AVTSTPGSKGVSASGGSGSVASGGSGSVASGGSGN--SRRTN-----	112			
DB	57	EITP-----KG--ASAQSGASQAQSGASQAQSGTSGPSGSGTSPSRSNTLPSN	109			
QY	113	-----PSDNSSDSKASYADLKHVRNYLLTTLKELYPOLFDTLNHMLTLCDNIHGFKY	166			
DB	110	TSSGASPPADASDSKASYADLKHVRNYLLTTLKELYPESLDLPNHMLTLCDNIHGFKY	169			
QY	167	LIDGYEINELLYKLNFFYFDLLRAKLNDCVANDYCOIPFNLIKIRANELDVLKLVFGYRK	226			
DB	170	LIDGYEINELLYKLNFFYFDLLRAKLNDCVANDYCOIPFNLIKIRANELDVLKLVFGYRK	229			
QY	227	PLDNKDNVGMEDYIKKNKTTIENINELTIESKTTIDKNKNA7KEBKKL	278			
DB	230	PLDNKDNVGMEDYIKKNKTTIENINELTIESKTTIDQNKNADNEBCKKI	281			
RESULT	11					
YD86	SCHPO					
ID	YD86_SCHPO	STANDARD;	PRT;	1957	AA.	
AC	Q10411;					
DC	01-OCT-1996	(Rel. 34, Created)				
DT	01-OCT-1996	(Rel. 34, Last sequence update)				
DT	15-JUN-2002	(Rel. 41, Last annotation update)				
DE	Hypothetical protein C1F3.06c	in chromosome 1.				
GN	SPAC1F3.06C					
OS	Schizosaccharomyces pombe (Fission yeast).					
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;					
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;					
OC	Schizosaccharomycetes.					
NCBI	NCBI_TaxID=4896;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=972;					
RX	MEDLINE=21848401; PubMed=11859360;					
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,					
RA	Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,					
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,					
RA	Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,					
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,					
RA	Holroyd S., Horsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,					
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,					
RA	Mooney P., Moulis S., Mungall K., Murphy L., Niblett D., Odell C.,					
RA	Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,					
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,					
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,					
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,					
RA	Woodward J., Volkheart G., Aert R., Robben J., Grymonprez B.,					
RA	Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,					
RA	Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,					
RA	Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,					
RA	Eger P., Zimmermann W., Wedter H., Wambutt R., Purnelle B.,					
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,					
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,					
RA	Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,					
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,					
RA	Domiguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,					
RA	Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,					
RA	Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;					
RT	The genome sequence of Schizosaccharomyces pombe."					
RL	Nature 415:871-880(2002).					
CC	CC					
CC	This SWISS-PROT entry is copyright It is produced through a collaboration					

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
CC

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; 270690; CAA94624.1; -
KW Hypothetical protein.
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 6.0%; Score 501.5; DB 1; Length 1957;
Best Local Similarity 22.2%; Pred. No. 1.5e-09;
Matches 352; Conservative 253; Mismatches 617; Indels 361; Gaps 67;

QY 105 SGNRRTPNSDSSDAKYADLKHVRNYLLTITELKYPQLFDLTNHLMTLCDNIHGF 164
DB 105 SGNRRTPNSDSSDAKYADLKHVRNYLLTITELKYPQLFDLTNHLMTLCDNIHGF 164
DB 162 SSKKKDKNTSVTLTSEEDVSYFQKKLTN-MESNFSAKOSEAYDLSRQLLTATEKDKK 220
QY 165 KYLDIGYEETNELLYKLNIFYDLIRAKLNDVCANDYCOIPFNKIRANELDVLKLVFGY 224
DB 165 KYLDIGYEETNELLYKLNIFYDLIRAKLNDVCANDYCOIPFNKIRANELDVLKLVFGY 224
DB 221 E--KDYETIKEDV-----SSIKASLAEEQASN-----KSLRGEQERLEKLLVSS 262
QY 225 RKLPIKDNKGVGMEDYIKKKTKTININELIEESKTTIDKNKATKEEKKLYQAQYD 284
DB 225 RKLPIKDNKGVGMEDYIKKKTKTININELIEESKTTIDKNKATKEEKKLYQAQYD 284
DB 263 NKTVSTLR-----QTSNRAECKTLQ---EKLE-----KCAINEEDSKLLEELKHN 306
QY 285 LSYNKOLEAHNLISVLEKRI-----DTLK-KNENIKELLDKINEIKNPPANS 333
DB 285 LSYNKOLEAHNLISVLEKRI-----DTLK-KNENIKELLDKINEIKNPPANS 333
DB 307 VANYDAIVHKKLIEDLSTRISFDFNLKSERDTLSIKNEKLEKLL----- 352
QY 334 GNTPTNLLD---KNKKIEEKEIKEIATI-----KFNIDSLFTDPLEEYLYREKN 383
DB 334 GNTPTNLLD---KNKKIEEKEIKEIATI-----KFNIDSLFTDPLEEYLYREKN 383
DB 353 RNTIGSLKDSRTNSQLEEBWELVESNRTIHSQLTDAESKLSFQENKSLGSDIEYQ 412
QY 384 KNTDISAKVETKESTNEPVGVTPLSYNDINNALNELNSFGDL-----INPFDYTKEP 439
DB 384 KNTDISAKVETKESTNEPVGVTPLSYNDINNALNELNSFGDL-----INPFDYTKEP 439
DB 413 NNLSSDKMKVQVSSQLEEARSLAH-----ATGKLAETNSRDFQNKIKDFEIQD 466
QY 440 SKNIYDNERKKFTNEIKETKEKK-----IESDKSYEDRSKSLNDITKEY- 488
DB 440 SKNIYDNERKKFTNEIKETKEKK-----IESDKSYEDRSKSLNDITKEY- 488
DB 467 LRACLNSS-----NELKESALIDKQDLNLRQIKQKVKSESTQSSLSQDLORDIL 521
QY 489 -EKLNIIDYSKNN-----NIDLTFKMMGKRYKYKVEK-----LTHNTFASYENSKHN 539
DB 489 -EKLNIIDYSKNN-----NIDLTFKMMGKRYKYKVEK-----LTHNTFASYENSKHN 539
DB 522 NKKKHEVYESQLNELKGELOTEISNSELSQLTLAAEKAATAVATNNELSESQSLQ 581
QY 540 L-----EKLTKALKYMEDYSLRNVVEKELKYKNLISKIENELETIVENIKKDEQL-- 592
DB 540 L-----EKLTKALKYMEDYSLRNVVEKELKYKNLISKIENELETIVENIKKDEQL-- 592
DB 582 LCNAFOEKLAKSVMLKENQNFSSLDTSFKKLNESHQELENNHQTITKOLKDTSSKLOQ 641
QY 593 -----FEKK-----ITKDNKPKDEKILEVSDIVKVOVQVLLMKNKIDELKKTOLILKNVE 642
DB 593 -----FEKK-----ITKDNKPKDEKILEVSDIVKVOVQVLLMKNKIDELKKTOLILKNVE 642
DB 642 LQLERANFEQKESTLSDENNDRTLKLLKLEESNKLKK-----QEDVDSLEKNTQTLKE-D 697
QY 643 LKNIHVNPYKQENKQEPYLLVLKKEIDKLKVPKPVKESLNEEKNNIKTEGQSD--N 700
DB 643 LKNIHVNPYKQENKQEPYLLVLKKEIDKLKVPKPVKESLNEEKNNIKTEGQSD--N 700
DB 698 LRKSEALREFSKLEAKN-----LREVIDNLK---GKHETL--EAORNDLHSSLSDAKN 745
QY 701 SEPTGEITGOATTPGQAGSALGSDSVQAQAEQKQAPPPVPPVPAKAQVTPPPA 760
DB 701 SEPTGEITGOATTPGQAGSALGSDSVQAQAEQKQAPPPVPPVPAKAQVTPPPA 760
DB 746 TNAILSELP-----KSESDVKRLTANVETLTQDSKAM----- 778
QY 761 PVNKNTEVSKLDYLEKLYEFLNTSYCHKYILVSHSTNMNEKILKQYKTKKEESKL-SS 819
DB 761 PVNKNTEVSKLDYLEKLYEFLNTSYCHKYILVSHSTNMNEKILKQYKTKKEESKL-SS 819
DB 779 -KOSFTSLVNSYOSINLYHELDRDH-----VNNQSQNNNTLL-----ESESKLTD 823
QY 820 CDPL-----DLLFNION-----NIPWYSMFDLSLNSLSQLPMWEYKEMVCNLYKLKD 868
DB 820 CDPL-----DLLFNION-----NIPWYSMFDLSLNSLSQLPMWEYKEMVCNLYKLKD 868
DB 824 CENLTQONTMTIDNVQKLMKHVHNOESKVSSELKEVNGKLS---LDLKNLRSSLNV-AISD 879
QY 869 NDKIKNLEAKKVSYSVKTLSLSSSQPLSLTPQDREYSANDTSH--STNLLNSIKLKF 926
DB 869 NDKIKNLEAKKVSYSVKTLSLSSSQPLSLTPQDREYSANDTSH--STNLLNSIKLKF 926

DB 880 NQILTLQALBSKNYDSLEQESQAQLNSGLSKLSBAEQKOLLHTENEELHRLDKITGLKIE 939
QY 927 ENILS--LGNKNKIYQELIGOKSSSENFYEKTLKDSDFYNESEFTNFVKSADDINSLNDES 985
DB 940 ESKSDOLGKGLTKARQEISNLKEENMSQ-----SOAITTS-VKSKLDE-----TLS 983
QY 986 KRKKLEEDINKLTKTLQSLFDLYNKYKLLKLERLFRDKKKTGVK-----YKMQIKKLTLLKEQ 1041
DB 984 KSSKLEADIEHLKKNKYEVEVERNALLASNERLMDLLKNGENIASLOTIEKKRAENDD 1043
QY 1042 LESKLSLNNPKHVLFQNFVFFNKKKEAEIAETENTLENTKILLKHKGLVLYKNYNESSP 1101
DB 1044 LQSKLWVSVSEYENLLLISSQTNKLSLEDTNQLKYTEKNVQKLLDEKQDNVELEELTSK 1103
QY 1102 LKLTSESIQTDENYASL--ENFKVLSKLEGLKLDNLEKKLSYLSLSSGHHLLIAELKEV 1160
DB 1104 YGLGGENAQIKDELLALRRKSKKQHDLCANFVDDDLKEKSDALEQLTNEKNEILVSLQ- 1162
QY 1161 IKKNYTGNSPSENNTDVNNALSEYKFKFLPEGTDTAVTVVSESGSDTLEQSQPKKPASTHV 1220
DB 1163 ---SNSNEALVEERSDLANRLSDMKKSLSDSDNVISVIR---SDLV----- 1203
QY 1221 GAESNTITTSQNVDDVDV-----IIVPIFGSEBEDYDDLGQVVTG----- 1262
DB 1204 -----RYNDELTLKKDKDSLSTQYSEVQCDRDDLLDLSLKGCEESFNKYAVSLR 1252
QY 1263 EAVTPSVIDNILSKIENEYEVLYKPLAGVYRSLKQLENNVMTFN-VNVKDI-LNSRFN 1320
DB 1253 ELCTKSEIDVPVSEILDDNFVFNAGNFSRLTVLSLENYLDAFNQVFNKKMELDNRLT 1312
QY 1321 KRE-NFKNVL-----ESD--LIPYKDLTSSNYVVKDYPK-FLNKE----- 1356
DB 1313 TTDAEFTKVVADLEKLQHEHDDWLIQRGDLEKA---LKDSEKNFLRKEAEMTENIHSLEE 1369
QY 1357 -----KRDKFLSS-----YNYIKDSID---TDINFANDVLGYKIL----- 1389
DB 1370 GKEETKKEIAELSSRLDNLQATNKLKNQDLHLNKEIRLKEVDLKEKESLIISLEESLN 1429
QY 1390 -SEKYKSDLDS---IKKYINDKQGENEKYLPFLNNIETLYKTVDKI-----DL 1434
DB 1430 QKQKESLLDAKNEHMLDDTSRKNS---LMKEITSNSSLDKSFELASAVEKLGAL 1486
QY 1435 FVHLEA-----KVLNYYEYSNV-EVKIKELNYLTIQDKLADFKNNNFVGIADL 1485
DB 1487 QKIHSELSLIMENIKSQLQEAKEIKQVDESTIOELHEITASKNNYEGKLNKDXDIIRD 1546
QY 1486 STDYNH-NLL-----TKFLST 1501
DB 1547 SENIEQLNLLAEKSAVKRLST 1569

RESULT 12

USOL YEAST	STANDARD;	PRT;	1790 AA.
ID	USOL YEAST		
AC	P25386;		
DT	01-MAY-1992 (Rel. 22, Created)		
DT	01-MAY-1992 (Rel. 22, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Intracellular protein transport protein USOL.		
GN	USOL OR INT1 OR YDL058W.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-X2180-1A;		
RX	MEDLINE-91185402; PubMed-2010462;		
RA	Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,		
RA	Yamasaki M.;		
RT	"A cytoskeleton-related gene, usol, is required for intracellular		
RT	protein transport in Saccharomyces cerevisiae."		
RL	J. Cell Biol. 113:245-260(1991).		


```
QY 1330 ESDLPYKDLTSSVYVVKDPYKFLNKKRDKFLSSVNYIKDSITDINFA-----NDVLGY 1385
Db 1645 -----QELDSTQQAQK-----SBEERRAEVRKFOVEKSQLDSQAMLETKYNDLVN- 1691
QY 1386 YKILSEKYSKDLSDSIKKYINDKOGENEKYLPLFNNTIE---TLKYTN-----DKIDLFW 1436
Db 1692 ---KEQAWKRDEDTVTKRTDSDQOEIEKAKELDNLKAENSKILKEANEDEIDDLMLLV 1748
QY 1437 IHLEAKVLYNYTEKSNVEVKI 1457
Db 1749 TDLDERNAKYSRKLKDLGVEI 1769

RESULT 13
RBPL_PLAVB
ID RBPL_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBPL.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Gallinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites";
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M68097; AAA29743.1; -
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 5.7%; Score 477; DB 1; Length 2869;
Best Local Similarity 20.9%; Pred. No. 1.4e-08;
Matches 372; Conservative 300; Mismatches 617; Indels 490; Gaps 87;

QY 115 DNSSDSDAKSYADLKURVRNYLTITKELYPQLF-DLTNHLMLFCDNIHGFKYLIDGYEE 173
Db 938 DQNEVSTAKA---LKEKIVSDSLRDKIDQYETFEKETSVAENTVSTIOSLSKAIDSLKR 994
QY 174 INELL---YKLNFFYDLIRAKL-----NDVCANDYCOI-----PFN 206
Db 995 LNSGINCKRYNTFDIILRSKIKTLREEVQKEMPKRGDKCGENTTALLKLSLRDKMGKIN 1054
QY 207 LKI---RANEDVLKK-LVFGYRKPLDNI---KDNVGKME-----DYIKKNKKTINEN- 253
Db 1055 EKLNDGRNLSLDTFKEDLLKFYSESKSIHLSKDKQGPDPPLNRIDEDWEDIKRDVDELNV 1114
QY 254 --ELIBESKKTIDKNNATKEEKKLYQAQYDLSTYNKQLEBAHNLISVLEKRIDTLKK 311
```

```
Db 1115 NYQVISENKVTLFPKNNSVT-----YIEAMHSHINTVAHGI-TSNK 1153
QY 312 NE---NIKELLDKIN-----EIKNPPANSNGTPTNTLLDNKK-IEEHEKEKEI- 357
Db 1154 NEILKSVKEVEDKLNVLVEONEDYKKVKNPENENKQLEBAIRGSMKLEKVINKHVSEMTQLE 1213
QY 358 --AKTKFNIDSLFTDPLELEYLYREKNKN-----IDISAKV-----ETKESTEPNEYPNG 406
Db 1214 STANTLKSNAKG-----KENEHDLLELNKTKGOMRDIYEKLLKIAELKEGTV----- 1261
QY 407 VTYPLSYNDINNALNELNSFGDLINPFDTYKPESKNIYTDNERKKFINELKEKIEKKEKK 466
Db 1262 -----NELKDANEKANKVEP-----EPERNI-----IGHLEIRITVEKDK 1296
QY 467 IESDKKSYEDRSKSLNDITKEYEKLLEIYDSKFNENNIDLTNFEKMMGKRYSK-VEKLT 525
Db 1297 A---GKVEE-----MNSLTKIEKLIQETSDS-QNELVTTSTIKHLENAGVEDVIKRN 1348
QY 526 HNTFASYENSKHNLEKLTAKALYMEDYSLR-----NIVVEKELKYKNLIS-KIENE 577
Db 1349 EEDSIOLREKAK-SLETLDMMKKLVQOVNMLQSAIQGNAGISKELNELKGVIELLISTN 1407
QY 578 IETLVENIKKDE-----BQLFEKKITK--DENKPEKILEVSDIVKQVQKVLNNKID 629
Db 1408 YSSILEYVVKKNSSSEVRFSQLANGEFTKABGEENASARLAEAEKLEKQIVKDOLDYSID 1467
QY 630 E-LKLTQTLILKNV-ELKHNI-----HVPNSYKQENKQEPYIL----- 664
Db 1468 DKVKKTEGIEKRELKMKESALTPEESEXFKQMCSSHMENA--KEGKKLEYLKNNGDGG 1525
QY 665 -----IVLKEIDKLKVF-----MPKVESLINE----- 687
Db 1526 KANITDSQMEEVGNYSKAEHAPHTVEAQVDKTKAFCEISIVAYVTRMDNLFNSLMKEVK 1585
QY 688 ---EKKNIKTEGOSDNSESTEGEITGQATTKPGQAGSAGEGDSVQAQAEQKQAPPV 744
Db 1586 VKCEKNDKAEKYSALKLP-YDGRI--KARVSENERKISEL-----KEKAKVEKKESSQL 1637
QY 745 PVPPEAKQAQVPPPPAPVANKTENSVKLDYLEKLYEFLNTSYICHRYILVSHSTMKEIL 804
Db 1638 NDVSTKSLLOIDNCRQQLDSVLSNIGRVK--QNALQYFDSADKSMKSVLPISLGAEKSL 1695
QY 805 KQYKITEESKLUSSCDPLDLFNIONNPVMTSMFDSLNNLSQLFMEIYEKEMVONLY 864
Db 1696 DKVKAAKESYEK-----NLETQONEM-----SRINVEGSLTDIDKKTIDTEN--DLL 1741
QY 865 KKK---DNDKIKNLLEEAKKVSSTVKTLSSSMQPLSLTPQDKPEVSANDDTSHSTNLNN 921
Db 1742 KMKQYEEGLLQKIKENADKRKSNFELVGS-----EINALDPSSTIFIKL 1787
QY 922 SLKLFENILSLGKNKI-YOELIGO-KSENFYEKILKDSDFYNESTFNFKVSKADDIN 979
Db 1788 KLKEYDWTGDL-KNYGVKMEIHGEFTKSYNLIETHLSNA-TDYSVTF-----EKAQSLR 1840
QY 980 SL---NDESKRKKLEEDI---NKLKK--TLQLSFDLYNKYKLLERLFDKKTVTGVKMQ 1031
Db 1841 ELAAKEEHLRRREEAIFLLNDIKKVESLKLKEMMKVSAEYEGMKRDHTSVSQLVQD 1900
QY 1032 IKKLTLLKQESKLSLNNPKHVLQNFVFFFNKKKEAEIAE---TENTLENTKILLKHY 1088
Db 1901 MKTIV---DELKT-LNDISECSSVLNNVSVIVKKVKESKHADYRRDANGSYESMVT--- 1952
QY 1089 KGLVKYNYGESSPLKTLSESIQTEDNYAS---LENFKVLSKLEGLKDLNLEKKKLSY 1145
Db 1953 --JANTFLSDEAKISSGMEFNAEMKSNFKTDLELEIFSVIS-----NSNELKKIEQ 2002
QY 1146 LSSGLHLLIAELKEVINKNKNTGNSPENNTDVNNALESYKKF---LPEGTDVATVSVES 1202
Db 2003 DSDNDVIQKRESQLAQDA-----TDIYNVILKNEPNEKLEEAKEEVVSEK 2051
QY 1203 GSDTL-----EQSQPKPASTHVGAESENTTITTSQNVDEVD 1238
```


Db 1005 QNNKEYEYKSEYKSLQNDLDOQTIYANTAOQNNYQEQELQHADVSKTISELREQ 1064
QY 855 YEMWCVNLYKLDNDKIKNLEAKKVSSTSVK-----TLSSSMOPLS-----LTPQ 902
Db 1065 HTYKGQVKTUWL-SRDOELNALKENKSWSSQESLLEQLDLSNRTEDUSSQNKLLYDQ 1123
QY 903 DKPEVSANDTSHSTN---LNNSKLFPENTLSLGNKNYI--QELIGOKSENFEYKI-- 955
Db 1124 IQIYTAADKEVNSTNGPGLNNIL-----ITLRREDILDTKVYVAERDAKMLRQKISL 1177
QY 956 ----LKSDDFYNESFT-----NFVKSADDINSL-----NDESK 987
Db 1178 MDVELODARTKLDNSRVEKENHSHSIIQQHDDIMEKLNQLNLLRESNITTLRNELENNNNK 1237
QY 988 KLEEDINKLKKTL-----OLSPFLYLNK-YKLKLERLFDKKTGVKYMOKIKLTL 1038
Db 1238 KEIQSELDKLQNVAPIESLTKALYSMOEBOELKL-----AKEVHRWKKRSQDILEK 1292
QY 1039 KEQLES-----KLNS-LNNPKHVLQN-----FSVFNK-KKEAE-----IAE 1073
Db 1293 HEQLSSDYKLESEITENLKEELENKROGAEEKFNRLRQAEKRLKTSKLSQDSLTE 1352
QY 1074 TENTLENTKLLKHKGVLKYNGESSPLKTLSESIQTDENYASLENFKVL---SKLEG 1130
Db 1353 QVNSLRDAKNVLEN-----SLSEANARIEE---LQNAKVAQGNQOLEA 1392
QY 1131 --KLKDN-----LNLEKKLSYLS--SGLHHLIAELKEVIK-----NKVYTGNSPS 1172
Db 1393 IRKLQDAEAKSRELQAKLEBSTSTVESTINGLNEELTLKELEKQROTOOQLOQATSAN 1452
QY 1173 ENNTDVNNALESYKFLPECTDVATVYVSESGDTE---OSQPKPASTHVGAESNTITT 1229
Db 1453 EQN-DLSNIVESMKKSFEE--DKIKFIKEKTOEVNEKILEAQERLQNSPINNEEIKKKW 1509
QY 1230 SQNVDDVEDVVIIVPIFGESEEDYDLGVQVVTGEAVTPSDINILSKIENEYEVLYKPL 1289
Db 1510 ESEHEQVSOIKI-----REAEALAKRIRLPTTEKIN-KIIEKKEKELEKEFEKEVEERI 1563
QY 1290 AGVRS-----LKKOLENNVMTFNVVKDILNSRKNRENFKNVLESDLIPYKDLTSSN 1343
Db 1564 KSEQSGEIDVLRKQLEAKVQ-----EKQKELENEYNKK-----LQEL---KDVPHSS 1610
QY 1344 YVVKDPYKFLUNKRDRKFLSSYNIKDSIDTINFANDVLGYKILSEKYSKLSLSIKKY 1403
Db 1611 HISDD-----ERDK-----LRAEIESR-----LREEFNELQAIKKK 1642
QY 1404 INDKQGENEYKLPFLNNIETLYKTVDNKIDLFIHLEAKVLNTYEKSNVEVKIKELNYL 1463
Db 1643 SFD-EGKQ-----QAMMKTT-----LLERK-----L 1662
QY 1464 KTODKLADPKKNNFVGIADLSTDYNHNNLLTKELSTGMVFENLAKTVLSNLLDGNLQ 1523
Db 1663 AKNESQLSETKQSAE-----SPPKSVNVQNPGLGLPRKIEENSFPNPLSG----- 1711
QY 1524 MLNISQHCQVKQCPQNSGCFRHLDEREECKCLLNTKQEGDKCVENPNP 1572
Db 1712 -----EKLKLNKSSSGGFNFTSPSPNKHQNDNDKRESLANKTDP 1754
RESULT 15
RBP2_PLAVB
ID RBP2_PLAVB STANDARD; PRT; 1251 AA.
AC Q00799;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 2 (Fragment).
GN RBP2.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88098; AAA29744.1; -.
KW Malaria; Receptor; Membrane.
FT NON_TER 1
FT NON_TER 1251
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;
Query Match 5.4%; Score 457.5; DB 1; Length 1251;
Best Local Similarity 21.1%; Pred. No. 2.5e-08;
Matches 306; Conservative 228; Mismatches 458; Indels 455; Gaps 71;
QY 272 EEEKKLYQAQYDLSYNNQLEEAHNLISVLEKRIIDLTKKNENIKELLDKINIKPPPA 331
Db 1 EFDKEKVKDTSDEK--KKSIKAY-----EKMGTLEK-----LEKMDDEKN--- 41
QY 332 NSGNTENTLLDNKKIEEHEKEIKETAKTIKFNID---SLETPDLELEYLYREK---NK 384
Db 42 -----IKEVEEAQIQKRIE-----IDHVNLMNDEVEKSKIVMEKIELYKK 84
QY 385 NIDISAKVETKESTENEPNGVT-----YPLSYNDINNALNELNSFGDLINPDYTKEPS 440
Db 85 EID-----EIKQKT--NEYKQDTSNFYYTEQYNSATQSKAKIEQFINIATTKGSDTS 137
QY 441 KNIYTDNERKKFNEIKEKIEKKIESDKKSYEDRSK---SLNDITREYKLLN--EI 495
Db 138 QDI---NE-----LESKEEVHKNLQLVKQESNMEERKQILSMKDLL-----ILNSET 185
QY 496 YDSKFNNTDILTNFKEMGMKRYKVEKLTHTHTFASYENSKHNLEKTLKALYMEDYSL 555
Db 186 IAKEISNNT-----QNALGFRENAKTKLAKTDELQ----- 216
QY 556 RNVVEKELKYKNLISKIENEIETLVENIKKDEBQLEFKKTKDENKPKDILEVSDIV 615
Db 217 RVAAMTEEAKAHK-----NNIDIALEDAQID-----TEVSKIE 249
QY 616 KVQVQVLLMNKIDELKTKQLILKNVELKHNHVPNSYKQENKQEPYLLIVLAKEDK 675
Db 250 QINRE---TMNKDEIKSYLSEIKYKDKCTTEISNKRKGDK-----IEFLE 294
QY 676 VPMKVES-----LINEKKNIKTEGQ-----SDNSEPSTEGEITGOATTKPGQOAGSA 724
Db 295 KFKPNEESNKNVINEINENINRSQYLYKDIEDAKQASTKVELFKHET---TISNIF 351
QY 725 LEGDSVOAQAEQKQAPVPVPEAKAQVPTPPAPVNNKNTENVSKLDYLEKLYEFLNT 784
Db 352 KESEILGVETKSQKI-----NKAEDIMK--EIER----- 379
QY 785 SYICHYILVSHSTMNEKILKQYKIKTKBEESKLSKSCDPLDLFIQNNIPVMI--SMPD 843
Db 380 -----HNSEIQTQVKGFOENLKL-----NEPHNYDNAEDEL 411
QY 844 NNSLSQLEWIEYKEMVCNLYKLKDN-DKIKNLEEAKKVSTSVKTLSSSSMOPLSLTPQ 902
Db 412 NNDKSTNAKVLIE---TNLESVKHNLSEITNKGGEKIYKAKDI---MQIKATSE 463
QY 903 DKPEVS---ANDTSHSTNLLNLSKLFIENILSLGKNK-----NIYQELIGQKSEN 950


```
Db 464 NTAETKLEKVDQSNVYVYLAQITERNLIVTEKRNLRGIDSTTINIEGAL---KESKG 520
QY 951 FYE-----KILKSDTFYNESTFNEVSKADINDSLND-ESKR 987
Db 521 NYEIGFLEBEIGKRNKRLKVITDKTSINSTGVNFFNFDLNOYDFNKINDYENKM 580
QY 988 ----KLEEDINKLKTQLSFDL---YNKYK-LKLE-----RLFDDKKKTGVKYMQIK 1033
Db 581 GEIYNEFEGLNKISENLNASENTSDYNSAKTLRLLEAQEKVNLNLEEANKYLRDVK 640
QY 1034 KUTL-----LKEQLESKNSL-----NNPKHVQNFSVFFNKKKEAEIAE-- 1073
Db 641 KVESFRIFNMKESLD-KINEMIKKQLFVNEGHGVNQLVENIKELVDENLSDILKA 699
QY 1074 -----TENTLEN-TKILLKHVGLVKNYNGESSPLKTLRESIQTDNYASLENF 1122
Db 700 TKNEEBIQITHTLKNKAKTILGHVDTSKTVGKITPELALTE----- 744
QY 1123 KVLKLEGLKDKNLNLEKKKLSYLSGLHLIAELKEVIKNKNTGNSPSENNTDYN-NA 1181
Db 745 -LLGDAKLAQELKPEK-----NNVLETEMNSKN-----TNELDVHKNI 785
QY 1182 LSEYK---KFLPEGTDVATVSESG-----SDTLEQSPKPA-ST 1218
Db 786 QDAYKVALEILAHSDIEDTKQKDSKSLIEMGNQIYLVKVVLIQYKNKISSIKSKEEAVSV 845
QY 1219 HVG-----AESNTITTSQNVDEVDVLIIVPIFGESEEDYDGLQGVVTGEAVTPSVIDN 1272
Db 846 KIGNVSKKHSLSKITS---DKSYDNIITAI-----EKQTELQNLNRSFTQEK-TNINSOS 897
QY 1273 ILSKIENEYEVL--YLKPLAGVYRSLSKKOLEN--NYMTFNVANVKDILNFRNKRNFKNV 1328
Db 898 KLEIKTFESLKNALKTLEGEVNALKASSDNHHEVQSKSEPVNPAL-SEIEKEET---- 952
QY 1329 LESDLIPYKDLTSSNVVVDKPKFLNKRERDRKFLSSYNIKDSI-----DPTDI--NPNAN 1380
Db 953 -----DIDSLNATLDE---LLKGRKTCEVSRKLIKDTVTKEISDDTELINTIBK 999
QY 1381 DVLGYKILSEYK---SILDSIKKYINDKQENKYLFP-----LNNIETLYKTV 1428
Db 1000 NVKAYLAYTKKYEDTVQDVLNHEFNTQVSNHEPTNFDKSNKSSELTAKVDTSKTI 1059
QY 1429 NDKIDLVFIHLPAKVLNITYEKSNEVEKIKEL-----NYLKTIOQDLADFKNNNFVGIAD 1484
Db 1060 ISLKGVIIE-----VNENTEMNTIESSAKEIYALNELKNKKTSLNEIYQTSNEVKLQE 1114
QY 1485 LSTDYNNHLLTKFLSTGVNFENAKTVLSNL-----LDGNLQGLMNLISQH- 1530
Db 1115 MKSNAD-----KYIDVSKIFNTVLDPTQKSNIVTNOHSINNVKDKLKGKLQELIDADSSF 1168
QY 1531 --QCVKK 1535
Db 1169 TLESIKK 1175

RESULT 16
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
```

```
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis."
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor."
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1."
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 215005; CAA78727.1; -.
DR PIR; S28261; S28261.
DR HSSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR MIM; 117143; -.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8CB8 CRC64;

Query Match 5.3%; Score 443.5; DB 1; Length 2663;
Best Local Similarity 18.4%; Pred. No. 1.6e-07;
Matches 345; Conservative 349; Mismatches 617; Indels 559; Gaps 82;

QY 126 ADLKHVRNVLTLIKELK---YPQLFDLTNHLT-----LCDNIHGRKY 166
Db 415 AKRRKRVTWCLGKINKMKNSYADQFNPTNITTTKLSINLLRDEISVCSSESVFSN 474
QY 167 LIDGVEEI-----NELLYKLNFFYDL--LRALNDVCANDYQCI-----PFNLKIR-ANE 213
Db 475 TLDLSEIETWNPATKLLNQENIESELNSLRADYDNLVL-DYQLRTEKEEMELKLEKND 533
QY 214 LDV-----LKKLVFG---YRKPLDN-----IKDNVKG 237
Db 534 LDFEALERKTKKQDQEMQLIHEISNLKLVKHREYVNDQLENELSSKVELLREKEQIKK 593
QY 238 MEDYTKNKKTTENTIN-----ELIEESKK-----TID-KKNKATKEEKKKL 278
Db 594 LQEYIDSQK--LENIKMDLSYSLESTEDPKMQKTLFDAETVALDAKRESAFRSLENLEL 651
```


RL J. Bacteriol. 175:7918-7930(1993).

CC - FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH

CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS

CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW

CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS

CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY

CC SIMILARITY).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U39701; AAC71437.1; -

DR EMBL; U02165; AAD12447.1; -

DR TIGR; MG218; -

KW Cytoadherence; Structural protein; Coiled coil; Complete proteome.

FT DOMAIN 28 838

FT COILED COIL (POTENTIAL).

FT DOMAIN 914 1591

FT COILED COIL (POTENTIAL).

FT DOMAIN 1632 1723

FT COILED COIL (POTENTIAL).

FT DOMAIN 1777 1804

FT COILED COIL (POTENTIAL).

SQ SEQUENCE 1805 AA; 216252 MW; 11D0933AF173284FD CRC64;

Query Match 5.2%; Score 440; DB 1; Length 1805;

Best Local Similarity 19.7%; Pred. No. 1.4e-07;

Matches 334; Conservative 301; Mismatches 566; Indels 492; Gaps 79;

QY 133 RNYLTI-----KELKYPOLFDTNHLMLTCDNIH-----GPKYL 167

DB 190 KNYLNVLDQYLNELDQLENOKRLLSIEYENTYRELVSADNQLQVYENIDQNOIQFHQ 249

QY 168 IDGY-EEINELLYKLN-----VFOLLRKLDVNCANDYQIIPFNKIRANEL-DVILK 218

DB 250 YTYRDELSQLERKQLTKQELVDKESALRVKIDDA-----DFYNARLAELDDVAK 301

QY 219 KLVGYRKLPLONI-KDNVGMEDYIKKNNKTJENIELIE-----ESKTTIDKNKNATKE 272

DB 302 QLSF-----QDGITKONAHQVEDKLVALNKEKDRLENTQKEAFFNLQRSALIDINK----- 351

QY 273 EEKKLYAQVYDLSYINQLEBAHNLISVLEKRIIDLTKNE-NIKELLDKINEIKNPPPA 331

DB 352 -----LQOENELFAHLEHQOQEF--QKQSDSLKLETEYKALQHKINEPKN----- 397

QY 332 NSGNTPTLLDKNKKIEHEKEIKEIAKTIFKNIDSLFTDPLEYLYLRXNKNIDISAK 391

DB 398 ESATKSEELLNOERELFEKRE-----IDLTLTQ-ASLEYEHORESSQLLKDQK 445

QY 392 VETKESTEPNPGVTPYPLSYNDINNALNELNSFGDLINFPDYTKPSKNITYDNERK- 450

DB 446 NEVKHQFNLEY-----AKKELDKERNLL-----DOQKKV 475

QY 451 --KFTNEIKETIKIBKKIES---DKSYVEDRSKSLNITKEYEKLLEIVDSKFNND 505

DB 476 DSEALFOLKEKVAQERKELEELYVKKQKQOKE--NELL-FFEKQLKQ-HOAPFENELE 531

QY 506 LTNFEKMGKRYSYKVEKLTHTNTFASYSKNSKHNLKLTALKYMEDYSLRNIVVEKELK 565

DB 532 AKQELFEAK---HALER-----SPIKLEDEKDLN--TKAQOITANEFSS-----QLK 573

QY 566 YKNIISIKIENIEFTLVENIKKDEQLPEKFKITKDN-----KPEKILEYSD--- 613

DB 574 TDKSADFELMLQNEYENLOQEKQKLFQERTYFERNAAVLNRLQOQREELLOQKETILD 633

QY 614 -----IVKQVQVLLMKNKIDELKTKTLQILKLNVELKHNHVPNS 653

DB 634 QLTKSFEQERLINQREHVELVASVEKQEILGKIKQDFSQTSLSNASKNLAEREMAIKKE 693

QY 654 KOENQKQEPYLLVLVKKETDKLVFMPKVESL---INEKKNIKTGG-----SDNSEPS 704

DB 694 KEIATEKQOLL-----NDVNNAEVIOADLAQNLQSLNQRSELOQNAKQRIADFNHDSLKKL 749

QY 705 TEGEITGOATTKPGQOAGSALEGDSVQAQA-----QEOKQA-----QPPVPVPVPEA 751

DB 750 NEVELSLQKRLQELQTLLEANQKQHSYQOAYFEGELDKLNREKQAFNLNRKQTMEVDAL 809

QY 752 KAOVTPPPAPVNNKTENVSKLDYLEKLYEFLNTSYCH-----KYLL 793

DB 810 KORLSD-----KHOALNMQQAELDRKTHE-LNNAFLNHDADQKSLQDQATVKTQKLD 863

QY 794 VSHSTMNEK-----ILKQYKITEEESKLSSCD-PLDOLLFNION 831

DB 864 LERSALLEKQREFAENAVAGFRHWSNKTSQLQKIYELTKKQSEQOTQKTELKIAF---S 920

QY 832 NIPVMSFSDLSUNSLSQLMEIYKEMVNCNLYKIKNDNDKIKNLLBEAKKSVSTVLTSS 891

DB 921 DLQKDYQVFELQKD--QEFQRQIEAKQ-----RELDKL-----AEK-NNQVKLELD 962

QY 892 SSMQPLSLTPQDKPEVSAN-DDTSHSTNL-----NNSL-----KLFENILSLG---K 934

DB 963 NRFQALQNKQDQTVQAQLELEREHQHNLQEQAFNAQNESLLKQREQLTKKIQAFHYELK 1022

QY 935 NKNIYOELIGOKSSSENFVEKILKQSD-----TFYNESFTNFVKSRADDINSND----- 983

DB 1023 KRNQFLALGKRLFAKEQDQQRKQDEINWRFKQPEKEYTDFDEAKKRELEELKIRSL 1082

QY 984 -----ESKRKKLEED---INKLKTQLSFDLYNKKYKLLERLFDKXKTKVQKY---KMOI 1032

DB 1083 QSNVELERKREKLATDFTNLNKVQHNTQINRDLNS---QIROFLLEKRNKQFQSEANA 1139

QY 1033 KKLTLKLEQESKLSLNNPKHVLFQNFVFFNNKKKEAEIAETENT---LENTKILLKHY- 1088

DB 1140 KKAFLIK-RLRSFASNKLQKEALAIQLEFQDKRDEQOQKELQOATLQLEQKFEKQNF 1198

QY 1089 ----KGLV-----KYNGESSPLKTLSEESI-----QTFEDNVASLE 1120

DB 1199 TEKQRLVAIKTQCEKLSDEKALKQKLVLANLSQTYLANKNKAEYSOQLOQKYNLL 1258

QY 1121 NPK-----VLSKLEGKLDNLNLEKKLSYLSGLHHLJIAELKEVYKNNK 1164

DB 1259 DLKENLERTKDQDKKHSIFARLT-KFANDLRFKKQLLKAQRIVDKDNRLKERNL 1317

QY 1165 NYTGSPENNTDVNNALESYKFLPEGTDVATVVSSESDTLEQSQPKKPASTHVGAES 1224

DB 1318 HFLSNETERKRAVLEDQISYFEKQKQATD-AILASH-----KEYKKKEGELQKLLVEL 1370

QY 1225 NTITTSQNVDDVDVVIIVPIFGESEEDYD-----DLGOVVTGEAVTPSIDNLSKI 1277

DB 1371 ETRKTKLND-----FAKFSROREFEFNORLKLLELQKTLQOTQTSNNFKTKAIQEI 1422

QY 1278 ENY-----EVLYLKP-----LAGVYRSLKQLENNVMTFNVNKD-----ILNSR 1318

DB 1423 ENSYKRGMEELNFQKFKDNKSRLYEYFRKMRDEIRKESQVKLVLTETORKANLLEAQ 1482

QY 1319 FNKRNFKNVL---ESDLPIYK-----DLTSSNVVVKDYPKFLNKEK-----RDKPL 1362

DB 1483 ANKLIENTIDFKEKELKAFKDKVDQDIDSTNQRKELNELLNKLQOQLIERERAI 1542

QY 1363 SSYNYIKOS-----IDTDINFAND-----VLGYKYLSEKYKSDL-----DSIKK 1402

DB 1543 NS-----KOSLLNKKIETIKRQLHDKEMVRLVLDNRKLAEQYQYQYQYQYQYQYQYQY 1598

QY 1403 -----YINDKQEN-----ERYLPP-----LNN 1420

DB 1599 DIKNFPPLFKINGNDMAFPYLPWLYPOOKODNTLOIRQLFEQLOLQPMQOQYENELNE 1658

QY 1421 IETLYKTVNDKIDLEVIHLEAKVLNYYEKSNNVEKIKELNYLKTIQKLADEFKNNNFV 1480

DB 1659 LRRQRNLLEKKLD--QIOLESOLNNKQSEFSKVESMMEKL--LEKTESRLNDFDQKINY- 1713

QY 1481 GIADLSTDYNNHN 1493

DB 1714 ----LTKKVNQHN 1722

RESULT 18

GOG4.HUMAN STANDARD: PRT: 2230 AA.
ID GOG4.HUMAN Q13439; Q14436; Q13270; Q13654; PRT: 2230 AA.
AC Q13439; Q14436; Q13270; Q13654; PRT: 2230 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE golgi autoantigen, golgin subfamily A 4 (Trans-Golgi p230) (256 kDa
DE golgin) (Golgin-245) (72.1 protein).
GN GOLGA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96215236; PubMed=8626529;
RA Erlich R., Gleason P.A., Campbell P., Dietzsch E., Toh B.-H.;
RT "Molecular characterization of trans-Golgi p230: a human peripheral
RT membrane protein encoded by a gene on chromosome 6p12-22 contains
RT extensive coiled-coil alpha-helical domains and a granin motif";
RL J. Biol. Chem. 271:8328-8337(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 131-2230 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=96125112; PubMed=8537393;
RA Fritzler M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
RT "Molecular characterization of golgin-245, a novel Golgi complex
RT protein containing a granin signature";
RL J. Biol. Chem. 270:31262-31268(1995).
RN [4]
RP SEQUENCE OF 524-672 FROM N.A.
RC TISSUE=Gastric fundus;
RA Balague C.;
RL Thesis (1994), Instituto municipal de investigacion medica, Spain.
CC -!- FUNCTION: MAY PLAY A ROLE IN VESICULAR TRANSPORT FROM THE TRANS-
CC GOLGI.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE CYTOPLASMIC FACE OF THE
CC GOLGI MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms: 1 (shown here), 2, 3
CC and 4; are produced by alternative splicing.
CC -!- DISEASE: ANTIGEN IN THE AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME AND
CC IN HEPATITIS B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U41740; AAC50434.1; -
CC EMBL: X82834; CAAS8041.1; -
CC EMBL: U31906; AAC51791.1; -
CC EMBL: X76942; CAAS4261.1; -
CC Genew; HGNC:4427; GOLGA4.
CC MIM; 602509; -
CC InterPro; IPR000237; GRIP_domain.
CC Pfam; PF01465; GRIP; 1.
KW Golgi stack; Antigen; Coiled coil; Alternative splicing.
FT DOMAIN 133 237 COILED COIL (POTENTIAL).
FT DOMAIN 276 1011 COILED COIL (POTENTIAL).
FT DOMAIN 1033 1214 COILED COIL (POTENTIAL).
FT DOMAIN 1259 2152 COILED COIL (POTENTIAL).
FT VARSPLIC 2154 2185 TPYKGNLYHTDVSFLGPEYELRKVLFEY -> HLTKV
FT VARSPLIC 2154 2185 AICTIRMSHLENPLNSICEKFLSI (IN ISOFORM
FT 2).

FT VARSPLIC 2186 2230 MISSING (IN ISOFORM 2).
FT VARSPLIC 2103 2109 MISSING (IN ISOFORM 3).
FT VARSPLIC 2220 2230 FTSPRSGLF -> SWLRSSS (IN ISOFORM 4).
FT CONFLICT 188 188 R -> K (IN REF. 3).
FT CONFLICT 220 220 Y -> H (IN REF. 3).
FT CONFLICT 276 276 T -> A (IN REF. 3).
FT CONFLICT 584 584 K -> E (IN REF. 3).
FT CONFLICT 628 628 T -> A (IN REF. 3).
FT CONFLICT 630 630 K -> E (IN REF. 3).
FT CONFLICT 682 682 K -> N (IN REF. 3).
SQ SEQUENCE 2230 AA; 261139 MW; 3BB733DB1EA86134 CRC64;
Query Match 5.0%; Score 424; DB 1; Length 2230;
Best Local Similarity 19.4%; Pred. No. 5.8e-07;
Matches 343; Conservative 303; Mismatches 57; Indels 572; Gaps 81;
QY 26 QELVKKLEALE-----DAVLGTGSLFOKEKMWLNESGTAVTITSTPGSKGSVAS 75
DB 137 EQLIQLRLRMERSLSYRGKYSELVTAYOMLOREKKL-----QGLS- 179
QY 76 GSGGSGVASGSGVASGSGVASGSGGSGNRTPNSDSSSDSAKYADLKHVRNY 135
DB 180 -----QSQDKSLRRIRAELEELQMD 199
QY 136 LTIKELKYPQLFDLTNHLMTLCDNIHGFYLDIGVEEINELLYKLNFFDILLRAKLDV 195
DB 200 QQAKKHLQ--EEFDAS-----LEEKDQYISVLQTVSLKQLRNG 238
QY 196 CANDYCOIPFNKIRANELDVLKLL-----VFGYR-----KPLDN 230
DB 239 -----PMN-----VDVLKPLPOLEPAQEVFTKEENPESDGEVVDGTSVKLTET 283
QY 231 IKDNYGMEDYIKKKTKTIENINE---LIEESKTKIDKNKATKEP-EK-KKLYQAQYDL 285
DB 284 LQQRVKRQENLLKRCCKETIOSHKEOCTLTSEKALQEQDLERLOLEKIKDLHMAEKT 343
QY 286 SIYNKQLEEAHNLISVLE-----KRIDLK-KNEIKELDKKINEKKNPPAN 332
DB 344 LL--TQLRDAKLLIEOLEQODKGWIAETKQRMHETLEMKEEETIAQURSIKQM-----TT 396
QY 333 SGNTPTNLLDKNKK--IEEHEKEIKEIAKT-----IKFNIDSLF-----TDPLELE 376
DB 397 QGEELREQEKSEKRAFAFEELKALSTAQKTEEARRKLKAEMDEQIKTIEKTSSEERISLQ 456
QY 377 YYL-REKKNKIDISAK-----VETKESTEPENYNGVITYPLS 412
DB 457 QELSRVQKQVWDVMMKSSSEQIAKLQKHEKLARKEQELTKLQTREREFQOMKVALE 516
QY 413 YNDINNALNELSFGDLINFPDYTKEPSKNIYTDNERKKFNEIKKIKIEKKI--ESD 470
DB 517 -----KSQSEYLIKISQEKQOESLAELEELQKAILTESE 552
QY 471 KKSIEDRSKSLNDITKEYEKLLENIYDSFNNDITNFEKMMGKRYSYKVKELTH----526
DB 553 NK-----LRDLOQEAITYRILELE-----SSLEKSLOENKNOSKDLAVHLEAE 597
QY 527 ----HNTFASYENSKHLEKLTALKYMEDYSLRNIVVEKELKYKMLISKIENIEITLVE 593
DB 598 KKNHKEITVMVEKHKTE--LESLEKHQD--ALWTEKLQVLKQOYQTEMEKLEKCEQKE 654
QY 584 NIKKDEEQLEFKKITDENKPDKEILEVDIVKVOQVLLMKNKIDELAKTQ-----L 636
DB 655 TLTKDKEIIFQAH-----EEMNEKTLEKIDVQKOTELLES--LSSELSEVLKARHKLLEELS 708
QY 637 ILKNV--ELKHNHVPNSYKQENKQEPYVILVLKKEID-----KLKVPMPKVESLINEE 698
DB 709 VLKQDTDKMQEELAKMDEQKNHQQOVDLSIKEHEVSIQTEKALKDQINLELLKER 768
QY 689 KKNIKT-EGOSDNSEPS-----TEGEITGQATTKPGQOAGSALE-GDSVQAQAKQQAQPP 743
DB 769 DKHLKEHQAHVENLEADIKRSEGL-----QOASAKLDVFSQYQSATHEQTKAY--817
QY 744 VPVPPEAKAQVPTPPAPVNNKNTENVSKLDYLEKLYEFLNTSYICHKYILVSHSTWNEKI 803

Db 818 -----EQLAQLO--QKLLD-LET-----ERI 836
Qy 804 LKQYKITEEESKLSGCDPLDLFFIONNIPVMYSFDSLNLSLQ---LFWIYVEKEMV 860
Db 837 LTRQVAEVEAKQKDVCTELD-AHKIQ--VQDMQOLEKQSEMEQKYSKTQVYES--- 890
Qy 861 CNLYKLDNDK----IKNLEEAKKVSSTLTSSSQPISLTPQDPEVSANDTSH- 915
Db 891 ----KLEDGNKEQEQTKQILVEKENMILQMRGQKKEIILT-----QKLSAKEDSIHI 940
Qy 916 ----STNLNLSLKLFENILSLG-----ANKNYIQELIGOKSEN-----FY 952
Db 941 LNEEYETFKQKQKMEKVKQAKEMQETLKKLLDQBAKLKLENTALELSQKQKOFN 1000
Qy 953 EKIL---KDSDFYNESFTNFVSKADDINSINDESKKLEEDINKLKTLL-QLSFDLY 1008
Db 1001 AKMLEMAQANAGISDAVSRLTNQEQIESLT-EVHRRLNDVSIWEKKLQNOQAEIQ 1059
Qy 1009 NKYKLERLFDKKTGKTKYKMQI-----KLTLLKEQLESKLSLNNPKHVLIQ 1057
Db 1060 ETHEIQLQ--EKEQVLAELKQILLFCGEKEEMKKEITWLKEGVKQDITLNELOEQLK 1116
Qy 1058 NFSVFN-----KKKEAIAET--ENT-----LENTKILLKHVKGLVYKYN 1097
Db 1117 OKSAHVNSLAODETKLKAHLEKLEVDLNLKSLKENTFLQEQVLVELKMLAEEDKRVKSEL-- 1174
Qy 1098 ESSPLKTLSEB--SIQT--EDNYASLEN-----FKVLSK-----LEGK 1133
Db 1175 -TSKLTDTDFEQFSKSHKSNKSLDEKSLFEFKLSBELAQLDIOCKKTEALAEKTN 1233
Qy 1134 DNLNLEKKKLSVLSGLH-----LAEKLEKIKKNTGNPS 1172
Db 1234 ELINSSSKTNAISRISHCOHRTTKVKEALLIKTCTVSELEAQLRQITEON-TLNLSP 1292
Qy 1173 ENNT-----DVNNALBSYKFLPEGDVATVVSSEGSDFLEQSQPKPASTHVGAESNT 1228
Db 1293 QOATHOLEKENQIKSMK-----ADIESLVTKEALQEGNQOQAAAE--KESCITQ 1343
Qy 1229 TSONVDDDEVDDIIV-PFGESEEDYDLDQGVVTGEATPSPVIDNLSKIENEYEVLYK 1287
Db 1344 LKKEUSENINAVTMLKEELKKEKVEISSKQTDNLNQ-----LQNSISLSEKEAAISLR 1400
Qy 1288 PLAGYRSLKKOLENNV--MFFNVN-----KDLNRSFRNRE 1323
Db 1401 K--QYDEKCELLDQVODLSFKVDTLSEKISALEQVDDWSNKFSEWKKKAQSRFTQH 1457
Qy 1324 NFKNVLESDLIPYKDLTSSNVVVDYKFLNKRKDKFLSSYNYIKDSIDTDINFANDVL 1383
Db 1458 NTVRELQIQL----ELKSKEAYEKDEQINLLKEELDQONKRFCDLKGEMDD----- 1505
Qy 1384 GYKILSEKYSKSDLS-----IKKYINDKQGENEKY 1414
Db 1506 ---KSKMEKESNTELELKTOTARIMELEDHITOKTIFIESLNEVLKKNYNOOKDIEHEL 1562
Qy 1415 LPFLNNITLKYTNKDKIDLEVIHLEAKVL---NYTYE-KSNVEYKIKELNY----- 1462
Db 1563 VQKLQHQFEL---GEEDKDNVKEAEKILTLNENQVYSKAELETKKLEHVNLSVKSK 1618
Qy 1463 ---LKTIOD-----KLADFKK 1475
Db 1619 BEELKALDRLESBSAAKLAELKR 1642

RESULT 19

Y109_YEAST
ID Y109_YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hypothetical 195.1 kDa protein in DNA43-UB11 intergenic region.
GN Y11149C.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Barrill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Chercher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule T., Odeil C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DSC-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z38059; CAA86129.1; -;
DR PIR; S48385; S48385.
DR SGD; S0001411; MLP2.
KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 5.0%; Score 423; DB 1; Length 1679;
Best Local Similarity 19.7%; Pred. No. 4.6e-07;
Matches 335; Conservative 296; Mismatches 583; Indels 488; Gaps 78;

Qy 198 NDYCOIPNLKIRANELDVLKLVGYRK--PLDNIKDNVGMKME---DYIKKN-KKTEN 251
Db 6 SEFLNVPEE-SLQGVTPVPLRKL--YKIAKFERSEEVTKLVLDVEIKSYYSRISK 61
Qy 252 INELIEESKKTIDKNKATKEE-----EKKLYQAQYDLSIYNKOLEAHNLISVL 302
Db 62 LKQLLDES---SEQNTAKEELNGLKQDLNEERSRYREID--ALKQLHVSHEAMREV 115
Qy 303 --EKRI-----DTLKNENIKELDKINIKNPPAN----- 332
Db 116 NDEKRVKBEYDIWQRDQNDLNDLKNENKLLRKLKMEMENILQRCSNAISLQLKYD 175
Qy 333 -SGVTPNTLLDNKKIEE-----HEKEIKEIAKTIFNIDSIFTDPL 373
Db 176 TSVOEKELMLQSKLLIEEKLSFSKTLTEEVTKSSHVENLEEKLYQMSYVESFT--- 232
Qy 374 ELEYLREKKNIDISAKVETK-----ESTEPNEYPNGVTYPLSYNDINNA--- 419
Db 233 ---YKFLNLLNKKQLSQSVEEKVLEKMKLKDTSVEKAEFFSKEMTLQKMMDLRSQLS 289
Qy 420 -----LNEINS-----FGDLNPFDTYK---EPSKN----- 442
Db 290 LEKDCSLRAIEKNDNSCRNPEHTVDIDELDTKLRLKSKNECORLQNVMDCTKEEA 349
Qy 443 -----IYTDNE--RKKFINEIKEIKIEKK-----IESDKK-----SYEDRS 478
Db 350 TMTTSVAVSPTVKGKLSDFIKRLQIKERNQKQFOLQNOLEDFILELHKTPELISFKERT 409
Qy 479 KSLNDITKEYEKLLEIYDSKFNNDILT-----NFKMMGKRYSY-KVEKLTTH 526
Db 410 KSLHEHLKRSTELLETSLTTRKOREITSRLQKINGCEANIHSLVKORLDLARQVKLL 469
Qy 527 HNTFASYN-SKHNLEKLTALKYMEDYSLRN-----IVVEKELKYYKNLSKIENETET 580
Db 470 LNTSAIOETASPLSQDELISLRLKILESSNIVNENDSQAIITRLVFNQVNVNELQKNVL 529
Qy 581 L-----VENTKKDEEQLFEKKITKDNKPNDEKILEVSDI--VKVOQVKVLLMKNID- 629
Db 530 LNCIRILADKLENVEGKQDKTLQKVENOTIKEAKDAITELNINAKMETRINILLRERDS 589


```
QY 630 -----ELKKTOL-----ILKNVELKHNHVPS 652
Db 590 YKLLASTEENKANTSVTSMEAREKKIRELEAELSSSTKVENSATIONLREKLLIYKQ 649
QY 653 YQENKQEPYLLI-----VLKKEIDKLKVPFK-----VESLINEEKKNIKTE-GQS 698
Db 650 CKKKTTFLEFENFKGLAKEKERMLEAIDHLKAELEKQKSWPSPSYTHVEKERASTELSOS 709
QY 699 DNSEPTEGEI-----TGATTKPGQAGSALGSDVQAQOEOQAOQPPVPV 746
Db 710 RKIKSLEVEISKLKETASPTPTKESLTRDPEQC-----CHEKKELQ-----M 753
QY 747 PYPEAKAQVPTPPAPVNNKTNVSKLD-----YLEKLYEFNTSYICHKYILVSHSTM 799
Db 754 RLKSEI-----SHNENKMFSSKEGQYKAKIKLENNL----- 787
QY 800 NEKILQYKITKEESKLSLSCDPLDLFNIONNIPWYSMFDSLNLSLQLPWEIYEEKM 859
Db 788 -ERLSDLSQKIQEIESIRSCDKSQKWK-AQNTI-----DDTENMKMSLLTETLSNKET 838
QY 860 VCNLYKLKDNKDKIKNLEEAQKYSTSVKTL-----SSSSMOPLSLTPQDKPEVSANDDTSH 915
Db 839 TIE-----KLSSLENDLKLRTKFOYKELQNSDASTLEPTLRKELEIQIQLKD----- 890
QY 916 STNLSNKLFLFENILSLGNKNIYQELIGQ--KSSNFYEKI-LKSDTFYNESFTNFVK 972
Db 891 ---ANSIQIQAIEIIS--SNEALIELKNELAKTNEYDAKIELEKKEKWAREEDLSRLR 945
QY 973 SKADDINSNDESKR-----KLEEDINKLKTQLSPDLYNKYIKLERL----- 1018
Db 946 GEIGETRALQPKLKEGALHFVQOSEKLRNEVERIQKIE-----KIEKMTIVQ 994
QY 1019 FOKKKTGVGYKQIJK--LTLKLEQESKLSNPNKPHVLQNFVFFNKKKEAEIAETE 1075
Db 995 LCKKEMSQYSTMKENKDLSELVIRLEDAADC-----QAEITKIK 1036
QY 1076 NTLNFKILL-KHYGLVYNGESSPLKTLSEESTOTEDNVASLENFVKLSKLEKLD 1134
Db 1037 SLSYSAQDLDKHER---KWMEEKADYERELISNIEQTES--LRVENSVLIEKVDVTAAN 1091
QY 1135 NLNLEKKLSYSSGLHHLIAELK-----EVIKNKNTGNSPSENNTDNNNALES 1184
Db 1092 NGDKHLKLVLSFNLRHENSLETKLTCTCKRELAFVKOKN---DSLEKTIINDLQRTQTL 1148
QY 1185 YKFLPEGTDVATVWSESGSDTLEQSPKPKPASTHYGAESNTI--TTSQNVDEVDVVI 1242
Db 1149 SEK---EYQCSAVIIDEFDITKEVTQ-----VNILKENNALQKSLKNVTEKNE--I 1197
QY 1243 VPIGSEBEDYDLGQ--VVTGEAVTPSVIDNLSKIENEYEVLYKPLAGVYRSLKKQL 1300
Db 1198 YKOLNDRQBEISRLDLQTLQTEQV--SINSNKLIVYSESEMOCKQR-----YQDLSQOQ 1250
QY 1301 ENNVMTFNVVKDI-----LNSRFKNRFKNVLESDDLIPYKDLTSSNVVVKDPYK 1351
Db 1251 K-----DAQKDIEKLTNEISDLKGLKSSAENANADLENKF-----NKLKQOAEH 1295
QY 1352 FLNKKERDAF-----LSSYNIKDSITDITNFANDVLGYKILSEYKSDLSIKKYIND 1406
Db 1296 KLDASKKQAALTNELNELKAIKLEQDLHFEN-----AKVIDDTKLKAHLEQSEDVS 1350
QY 1407 KOGENEKYLPLFNNIETL-----YKTVNDKIDLFIHLEAKVLNLYTEKSN-VEVYKIE 1459
Db 1351 RDHEKDTYRTLMEEIESLAKELQIFKTANSSDAF-----EKLKVNKEKEKDRIDERTKE 1406
QY 1460 LNYLKTIDQKLADFKNNNFVGLADIJSTDYNNHNNLLTKFLSTGMVFENLAKTVLSNLLDG 1519
Db 1407 --FEKLQETL-----NKSTSSEAYSKD-----TETLKKEWLKEYEDE 1443
QY 1520 NLQGLNISOHQCVKQCPONSGCFRHLDERECKLLNFKQK-GDKCVENPNP-TCNEN 1577
Db 1444 TLRIKEAENLKKRLPSEERIQIISKRKE-----ELEEFERKLEKENAGSLTFLDN 1498
QY 1578 NG-GCDADAKCTEEDSGSNGKK 1598
```

```
Db 1499 KSGGDAEEELWNPSKGNR 1520
RESULT 20
MYSI_YEAST
ID MYSI_YEAST STANDARD; PRT: 1928 AA.
AC P08964;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-I isoform (Type II myosin).
GN MYO1 OR YHR023W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=91088308; PubMed=2263482;
RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;
RT "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide
RT sequence."
RL Nucleic Acids Res. 18:7147-7147(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RL Science 265:2077-2082(1994).
RN [3]
RP SEQUENCE OF 1-760 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=88111539; PubMed=3322809;
RA Watts F.Z., Shields G., Orr E.;
RT "The yeast MYO1 gene encoding a myosin-like protein required for cell
RT division."
RL EMBO J. 6:3499-3505(1987).
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X53947; CAA37894.1; -
CC EMBL; X06187; CAA29550.1; -
CC EMBL; U10399; AAB68872.1; -
CC PIR; S05806; S05806.
CC PIR; S12323; S12323.
CC PIR; S46773; S46773.
CC HSSP; P08799; 1MND.
CC SGD; S0001065; MYO1.
CC InterPro; IPR000048; IQ-region.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00063; myosin_head; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.
```


Db 1130 IKGDIDSLFGE-----NDMLKTAKESNEAEIDRLKOKLQRSIENAKKYS 1173
Qy 1013 LKLERL---FDKKTVGCKYM-QIKKLTLLKEQLESKL-----N 1047
Db 1174 DALDKLRPEYDRLOQLYREKIKQAEMLTQAOVODLESRLNQSRRELDAOTDKLIASEGDRN 1233
Qy 1048 SLNPNPKHVLQFNSVFFNKK-----KKAETAETENTLENTKILLKHKYGLVKY-YN 1096
Db 1234 ALRSEVEKIQHEVOFMREQLLRKTDEYOAAALSDLVNAHRTAEDGRVNAQVLAERKYEIN 1293
Qy 1097 GESPLKTLSESTQEDENYASLEN-----FKVLKSLKGLKDKMLNLEKKLSVLSGLH 1151
Db 1294 DLQSLDNAEQLYLTLOQNYAVAVENRDMLYDALRHSMDTRVTINRFIL-----GVD 1348
Qy 1152 HLIARLKVIRKKNVGTGNSPSENNTDVNNALESYKKFLPECTGVATVVSSEGSTDLQSQ 1211
Db 1349 ESMDEKKTIVLOTOKSPDGKSKERFDISDL-----DTNQLKLGRIEKELEKNE 1398
Qy 1212 -----PKKPASTHVGAE-ESNTITTSQNVDDVDVIVIFGESEEDYDDLQ-QVVT 1261
Db 1399 YRDALDRKKKSIKINKQETIFT--NIEDQLVDV-----EERKRTLEMLAS 1447
Qy 1262 GEAVTPSVIDNI-----LSKIENEY--EVLYLKPLAGVYRSLSKKQLNENNVMT 1306
Db 1448 AKQLRSQEEALKQORDEERSHMKLKIAPFEMEARKEAQLRQLNELVRLNLRKDLTAQGD 1507
Qy 1307 FNV-----NVKD-----ILNSRFKNFKNVLESDDLIPYKDLT 1340
Db 1508 LGVLHDHERYVYAKFHLESKLKQESQIRLLVANFETERNLSNEKVRDLASRLQOT 1567
Qy 1341 -SSNVYVVDKPKFLNK-----EKRDKFLSYNYIKDSIDT---DINFAN 1380
Db 1568 ESKNADMKEDNRLKKDLKASTNEAELRRTIDQNSRVVSDNQILKQLESQAQNDLSNAN 1627
Qy 1381 D-----VLGYKILSEKYSKDSLDI---KKYINDKQGENEKYLPFLNNIETLYK 1426
Db 1628 NRKQOLENELLVSELRLDKQFSDNANRIIDLQHLTDADENKKRNTNLSLE---K 1684
Qy 1427 TVNPKIDLVHLEAKVLNRYEKSNEVKE-----LNYLKTIQDKLADFKNNFVFG 1481
Db 1685 TVS-----QORTIETIRQQLSLALNERNTLONDLRLQRR----- 1720
Qy 1482 IADLSTVDYNNHNLTKFLSTGWFENLAKTVLSNLLDGNLQGMNLISQHCVKKQCPONS 1541
Db 1721 LARMETEKINN--DKYDELEKIRASLIKRI--ELLDEKRTMENILHETALQREAISS 1776
RESULT 22
MYS2_DICDI
ID MYS2_DICDI STANDARD; PRT; 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Levinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
Dictyostelium discoideum";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a

phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
Dictyostelium myosin heavy chain";
FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
discoideum complexed with MgADP.Befx and MgADP.ALf4-";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
truncated head of Dictyostelium discoideum myosin to 2.7-A
resolution";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
Dictyostelium discoideum myosin motor domain to 1.9-A resolution";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammAS, and MgAMPPNP complexes
of the Dictyostelium discoideum myosin motor domain";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
Dictyostelium discoideum myosin motor domain";
RL J. Mol. Biol. 274:394-407(1997).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -!- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
(MLC-2)
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CORTEX.
CC -!- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CVS AT THE SH-1
POSITION (688).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

Qy 1529 OHQCVKQCPQNSGFRHLDERECKCLLNYKQEGDKCVENPNPTCNENGGCCDADAKCT 1588
 Db 1908 -RTVEAEADSKS-----EAEQSKRLVELEEDAR-----RNQKEIDAKEI 1948
 Qy 1589 EEDSGSN 1595
 Db 1949 AEDAKSN 1955
 RESULT 23
 ALMI_SCHPO
 ID ALMI_SCHPO STANDARD; PRT; 1727 AA.
 AC Q9UTK5; Q13313; Q9UT8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Abnormal long morphology protein 1 (Sp8).
 GN ALMI OR SPAC1486.04C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN
 RC SEQUENCE FROM N.A.
 RP STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Manes P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe."
 RL Nature 415:871-880(2002).
 RN
 RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=972;
 RX MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The almi+ gene from Schizosaccharomyces pombe encodes a coiled-coil
 protein that associates with the medial region during mitosis."
 RL Mol. Gen. Genet. 262:921-930(2000).
 RN
 RP SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN=968 h90;
 RX MEDLINE=20223868; PubMed=10759889;
 RA Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
 RA Hiraoka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 fission yeast cells by the use of a GFP-fusion genomic DNA library."
 RL Genes Cells 5:169-190(2000).
 CC
 CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN

CC CYTOKINESIS.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL133357; CAB62414.1; -
 CC EMBL; AF010473; AAB65416.1; ALT_INIT.
 CC EMBL; AB028012; BAA87316.1; -
 CC Coiled coil.
 CC DOMAIN 57 361 COILED COIL (POTENTIAL).
 CC DOMAIN 443 463 COILED COIL (POTENTIAL).
 CC DOMAIN 542 740 COILED COIL (POTENTIAL).
 CC DOMAIN 804 1106 COILED COIL (POTENTIAL).
 CC DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 CC DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 CC DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 CC SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 DR
 DR EMBL; AL133357; CAB62414.1; -
 DR EMBL; AF010473; AAB65416.1; ALT_INIT.
 DR EMBL; AB028012; BAA87316.1; -
 KW Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 4.7%; Score 399.5; DB 1; Length 1727;
 Best Local Similarity 20.6%; Pred. No. 2.7e-06;
 Matches 336; Conservative 279; Mismatches 562; Indels 453; Gaps 80;
 Qy 231 IKDNVGMEDYIKRKKTKTINELIE-----ESKKTIDKNKATKE-----E 273
 Db 6 LEDDQLVHEFL--DVSFEDIKPLVSVNGFAVFAISAIKTKVKDINALKDQLVQEVNHE 62
 Qy 274 EK-----KKLYQAQVDLSYKNQKLEEAHNLISVLEKRTDTLKK-ENIKELDKNEIKN 327
 Db 63 HKNVLTKKINLEQLOQSSNNQAESRNLISVLNENESLKTLENQNKRFDA----- 117
 Qy 328 PPPANGTNPNTLDDK-NKITEEHEKEKEIAKTKFNIDSLFTDPLEYLYLRKKNNI 386
 Db 118 -----TTNQSLRANSELQEQSKIASQSLIAKDQIEAL-----QNSNS 157
 Qy 387 DISAKVETKESTEPENYGVYPLSYNDINNALNELNSFGDLINPFDYTKPSKNITYD 446
 Db 158 HLGEOVQS-----AHQALSDI----- 173
 Qy 447 NERKK-----FINEIKKIKIEK-----KIESDKSYEDRSKSL 481
 Db 174 PERKQHFASSSSRVKEILLVQEKASVLSALQSDHSKVCLEKLVSSROVQDEKKL 233
 Qy 482 NDITKEYEKLNE---IYDSKFNNDIDTNFEKMMG-----KRYSYKVEKLTHHTNPAS 532
 Db 234 AGLAQONTE-LNEKIQLEQKRSNYSSDGNISKILETDPTSIKELEEEVE--TKRLTAL 290
 Qy 533 YENSKHNLEKTKAL--KYMEDYSLRNIVVE-----KELYYKNLISKIENIEFTLVEN 584
 Db 291 WESKSELQSEVAALQEKLTQQSLYNNVTEELNKNKQOLLISENSLRELQEKYDSVYSE 350
 Qy 585 IKKDEEQ-----LFE---KKITDENKPKDEKILEV-SDIVKVOQVQVLLANKNK 630
 Db 351 LQVYKKNKNTSVSACVGLFSPLAQKLSAVQN-PESFTKTVYSDNNKLOQKVSLLKLQDR 409
 Qy 631 L-----KKTOLILKNVE---LKNHNVHPNS-----YKQENK 658
 Db 410 LTNKFSFCEQVKORIPVVKQORSEIVRNINYMFLSESLETSSNNLTQVQAEELLTKMR 469
 Qy 659 QEPYYL-----IVLKEIDKLKVPMPK-----VESLINEEKKNKTKTSGQ 697
 Db 470 QEACYQLQTASRTQCSDLSEVVICLMAELDHLNETKSRNPATVQVQVALDEYAQNESTA-- 527
 Qy 698 SDNESPSTGEITGATTKPG-----OAGSALLEGDSVOAQAQE-OKAQPPVPVPVP 749
 Db 528 ---SETLVNKELANFSSIKREAVSKTLELRKVRALCED-VEIQKQTVQVQISNAVKENSN 583
 Qy 750 EAKAQVTPPPAVN-NKTENVSKLDYLEKLYEFLNTSVYICHKYILVSHSTMNEKILKQYK 808


```
Db 584 TLSEQIKNLESELNKKIRKESLNLNRLNKKEMLATF-----RSSILSHNSSAGNI----- 634
QY 809 IYKEESKLSDDPLDLLENQNNIPVMYSMFDLSLNNLSQ-----LMEIYKEKWCN 862
Db 635 -----DOKKSID--ESTRELEKNEYEVYRNEMTAQESLKRQDLSLSEMAIRKLENS 687
QY 863 LYKLG-DNOKIKNLLLEAKVSTSVKTLSS--SSMOPLSLTPQDKPEVSANDDTSHSTNL 919
Db 688 KYQQQLSTDRLNANNDVFAFKKEALRSINQNLQDI-ISQDQRAKFAEELLHVNSL 746
QY 920 NNSLKLFENILSGK---NKNYIYQELIGQSSNFYEKIL-----KSDTFYNES 966
Db 747 AERLK-GELNASKGEXDLRKRTQERLI-----SEN--DKLLAERLMSLVSDQLQTLNQ- 798
QY 967 FTNFVSKADDINSLNDESKRKKLEEDINKLKKTLQ-LSFDLYNKYKLLERLFDKXKIV 1025
Db 799 -----QOLSDAARKVFESKESLSLSQKLKESNKMNDLHSLQK-----SLEKSG 846
QY 1026 GYKMKQIKKLTLLKEOL-----ESKLSLNNPKHV-LQNFVSF---FNKKK 1067
Db 847 IEYSSRIKTLMEKOSLSEDNKRLLDNQOMMEIKLOELNGVIELEKQRFSTLEAKTQOK 906
QY 1068 EBIATENTLEN--TKILLKHVGLVYNGESSPLKTLSESIOTEDNYA--SLENPK 1123
Db 907 NTSYSREALLSSLSLQSKH-----TSLESQYNSLRNIEQLQ 946
QY 1124 VLSKL-----EGKLDKNLLEKKLSYLSLSSGLHHLIAELKEVKN-----KN 1165
Db 947 AASKLAEMVVERKTYEYRITQTSLEKNNHLKITS--LEQRIVLQDEIASSSLRJCN 1004
QY 1166 YTGNSPS-----ENNTDYNNALESY---KFLPBGTDV---ATVVSSESGSDTLRQSOP 1212
Db 1005 IYKDSRTRVALLLEENKHLNLSHRNAEKQHLKENDYKQOOLLVTDLRKTRDYE- 1063
QY 1213 KRPASTHVAES-----NTITTSQNDDEVDVVIPIFGESE-----EYDDLQGV 1259
Db 1064 -KELLRHADARSTLQKREDYTKALEQVEDLNKEITAKAGINESQPPFISEKEDPLRQSV 1122
QY 1260 VTGEAVTPSVIDNLSKIENEVEVLKLP-LAGVYRSLKQLENNVMYTNVNVKDLNLSR 1318
Db 1123 YVLKKQONALLTQLOSSNLFNFAITSPSDLSVMKLGSLDQNHVVKRISKEMEIISCOR 1182
QY 1319 ----FNKREPNKVNLESLIPYKDLTSSNVVVKDPYKFLNKRKDKFLSSYNYIKDSIDT 1375
Db 1183 QLLFLENKKLKTVE-----SSNRVIAD-----LQGITKEDVSS---TSESVGER 1225
QY 1376 INFANDVLGYKTLSEKYKSDLDLSIKKYINDKQGENEKYLPPL-NNIETLYKTVDKIDL 1434
Db 1226 SNYLNWV-----ALLNESKSLRENLER-----NEEVITELREKIETL-----KTDL 1267
QY 1435 FVTHLEAKVLYNTYKESNVEVKIK-ELNLYLKTQDKLADFKKNNNFVGI---ADLSTDYN 1490
Db 1268 -----ANFLNKEQLESQLOTEKAAVKVLNSENEEKYKRNQELLSLNSSTSSD 1318
QY 1491 HNNLLKFLSTGVFNENLAKTVLSNLDGNLQGLMNI--SQHCQVKKQCPQNSGCFRHL 1548
Db 1319 ASRLKNELVSKNLEELAQEI-----GHLKSELETVKSSEDLENERAQNSKIEQLE 1372
QY 1549 ERE---ECKLLNYKQEGDKCVENPNPTCNENNGGDDADAKTEEDSGSGNGKKITC---- 1601
Db 1373 LKNTKLAAAWRTKYEQVNVKNSLEKHN-----QIRQOLSQRTSELEAKVAECHOJLN 1422
QY 1602 -ECTKPDSPY 1610
Db 1423 EQLNKPSATP 1432
RESULT 24
ID RA50_YEAST STANDARD; PRT: 1312 AA.
AC P12753;
DT 01-OCT-1989 (Rel. 12, Created)
```

```
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN DNA repair protein RAD50 (153 kDa protein).
OS RAD50 OR YNL250W OR N0872.
OC Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=REG21;
RC MEDLINE=89276917; PubMed=2659437;
RA Alani E., Subbiah S., Kleckner N.;
RT "The yeast RAD50 gene encodes a predicted 153-kD protein containing a
RT purine nucleotide-binding domain and two large heptad-repeat
RT regions.";
RT Genetics 122:47-57(1989).
RL [2]
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RC MEDLINE=97377992; PubMed=9234673;
RA Sen-Gupta M., Gueldeener U., Beinbauer J.D., Fiedler T.A.,
RA Hegemann J.H.;
RT "Sequence analysis of the 33 kb long region between ORC5 and SUI1
RT from the left arm of chromosome XIV from Saccharomyces cerevisiae.";
RL Yeast 13:849-860(1997).
CC -!- FUNCTION: INVOLVED IN DNA DOUBLE-STRAND BREAK REPAIR (DSBR). THE
CC RAD50/MRE11 COMPLEX POSSESSES SINGLE-STRAND ENDONUCLEASE ACTIVITY
CC AND ATP-DEPENDENT DOUBLE-STRAND-SPECIFIC EXONUCLEASE ACTIVITY.
CC RAD50 PROVIDES AN ATP-DEPENDENT CONTROL OF MRE11 BY UNWINDING
CC -!- SUBUNITING DNA ENDS INTO THE MRE11 ACTIVE SITE.
CC -!- SUBUNIT: FORMS A COMPLEX WITH MRE11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14814; CAA32919.1; -
DR EMBL; X96722; CAA65494.1; -
DR EMBL; Z71526; CAA96157.1; -
DR PIR; S05808; BWBYDL.
DR SGD; S0005194; RAD50.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004584; Rad50.
DR TIGRfams; TIGR00806; rad50; 1.
DR DNA repair; Hydrolase; ATP-binding; Coiled coil; Meiosis.
FT NP_BIND 34 41
FT DOMAIN 185 347
FT DOMAIN 403 558
FT DOMAIN 617 672
FT DOMAIN 734 1108
FT SEQUENCE 1312 AA; 152568 MW; 58A0AA173AC5677E CRC64;
SQ
Query Match 4.7%; Score 393; DB 1; Length 1312;
Best Local Similarity 19.8%; Pred. No. 3.3e-06;
Matches 264; Conservative 251; Mismatches 443; Indels 376; Gaps 60;
QY 204 PFNLKIRANELDVLKLVGYRKPLDNINKNYKQLEEAHNLISVLEKRIDTL-KNENIKELLDKI 263
Db 168 PSNLKKKDFEIQAMK---FTKALDNLKSIKKDMSVDIKLLKQSVLEHLK-----L 214
QY 264 DRKNATKEEEKKLYQAYQYDLSYNNKQLEEAHNLISVLEKRIDTL-KNENIKELLDKI 322
Db 215 DKDRS---KAMKLNTHQLQTKIDQYNEEYSEIESQLEITEKSKLKFKNQDFQKILSV 271
QY 323 NEIKNPPPPANS-----GNTPTNLLDKNKKTEEHEKEI---K 355
Db 272 ENLKNTKLSISQVQRKLSNIDILDSKPDQNLNLANFSKVLMDKNQRLDLETDISSUK 331
```



```
Query Match      4.6%; Score 386.5; DB 1; Length 2748;
Best Local Similarity 20.6%; Pred. No. 1.2e-05;
Matches 377; Conservative 300; Mismatches 640; Indels 513; Gaps 96;

QY 53 LNEGTSCTAVTSTPGSGSVASGGGSGSVASGVSAS-----GGSVASGG 98
DB 28 LSOEYSTNGVTRMANLKADECGSGDEGDKTRFSIISILSKRETKDVLPEFAGSSSHG 87
QY 99 SVASGGSGSRRT-NPSDN-----SSSDAKSYADLKHVRNVLLTIKELYPOLPD 149
DB 88 VLTANSKDMNTFLEUSENLVLCRKQLQSSNEAKN-----EQIKS-LKQIKESLSDKIEE 141
QY 150 LTNHMLTCLDNIHGFKYLIDYEEIN-ELLYKL-NFYFDLLRAK-LNDVCANDYCOIPEN 206
DB 142 LTNQK-----KSPMKELDSTKDLNWDLESKLTNLSMECROLKELKKYKTESWDEKES 194
QY 207 LKIRANELDVLKLVPG-----YRKPLDNTKQNGKMEYIKKNNKTIENINEL 255
DB 195 LKLLKTDLEILTNGMENDLSSQKLYDKETSELKE--RILDNNENDRLIISVSDL 251
QY 256 IEB-----SKKTIDKNK--NATKEEKKKLYQAY--DLSIYN-----K 290
DB 252 TSEINLSQNRTERIKIQLODDAKASISLARKVKQKYQKQHTSDTTVSDPDSEGT 311
QY 291 QLEEAHLISVLEKRTIDTLKKNENIKELDKINEIKNPPANGSNTPNTLLDNK---K 347
DB 312 SEEDIFDIEIDMETGSPVEDISEDLVKKYSEKNMILLNSDSYKNLLQKSESASKP 371
QY 348 EBEKEIKELIAKTIKNIDSLTDP--LELEYLRKKNIDISA-KVETKESTEP-NEY 403
DB 372 KODELTKVAENL--NMIALNDNDNYSKFEFSLESHIKYLEASGYKVLPLEEFENLN 428
QY 404 PNGVTPLSYNDINNALNELNFGDLINPDTKPEKSNLYTDNERK-----KFINE 455
DB 429 --SLSNP--SYNYLKEKLOAKIPIDOQSTFNLLKEPTIDFLPLTSKIDCLLIPTKDYND 485
QY 456 IKERIK-----IEKK-----KIESDKKSYE 475
DB 486 LFESVKNPSTEQMKCLEAKDLQSNICKWLEBRNGCKWLSNDLYPSWVKNKIETPSKYQ- 544
QY 476 DRKSLSNDITKEYELL--NEYIDSKFNNID-----LTNFEKMGKRYSYKV 521
DB 545 -----LSDRAKEYDQVLDITKALEGLKNPTIDFLREKASADYLLKKEDYVSPSLEYLV 599
QY 522 E--KLPHHTF--ASYEN--SKHNLEKLTALKYMEDYS--LRNIVBEKELKYKNLIS 572
DB 600 EHAKNTHHLLSDSAYEDLVCKENPD-----WEFLKEKSAKLGHTVVSNEA--YSELEK 552
QY 573 KIEN-EIETLVENIKKDEBQLFE---KKITKDENKDEKILEVSDIVKVOQVLLMKNK 627
DB 653 KLEQPSLEYLVHAHAKATNHLLSDSAYEDLVCKENPDMEFLKE----- 696
QY 628 IDELKKTLQILKNVELKHNHPNSYKQENKOEPPYLYLVLKKEIDKLVFMPKVESLINE 687
DB 697 -----KSALGHTVVSNEAYSELQK-----YSELEKEVEQ-----PSLAYLVH 736
QY 688 EKKN-----IKTEQSD--NSESTGEITGOATTPGQOAGSALGDSVQAOAQ--EQOK 739
DB 737 AKATDHLLSDSAYEDLVCKENPDVEFFLKEKSAKLGHTVVSSEYSELQRYSELEKEV 796
QY 740 APPVPVPVPEAKQVTPPPAPVNNKTNVSKLDYLEKLYEFINTSYICHYILVSHSTM 799
DB 797 EQPSLAYLVHAHA-----TDHLLSDSAYEELVCKENPDMEFLKEKSAKLGHTTV 848
QY 800 N-----EKILKQYKI-----TKREESKLSKCDPLDLLFNIIQNNIPVWYSMFDSLNS 846
DB 849 SNEAYSELEKKLEQPSLAYLVHAHAKATDHLLSDSAYEDLVCKENSDVEF-----LKEK 903
QY 847 LSOLFPEIYEKEMVNCNLYKLKNDKIKNLEAKKVVSTSVKTLSSSQMPLSITPDQKPE 906
DB 904 SAKLGHTVVSNEAYSELEKKLEQPSLAYLVHAHAKATDHLLSDSAYEDLVCKENSDVEF-----LKEK 952
```


RX MEDLINE-97105885; PubMed-8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkil E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE-97252497; PubMed-9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA Hermann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
RL pneumoniae cytoskeletal protein HMW2 and cytodherence";
RL J. Bacteriol. 179:2668-2677(1997).
CC - FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A000051; AAB96174.1; -
DR EMBL; U59896; AAB52527.1; -
DR PhosSite; P75471; -
KW Cytoadherence; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 31 880 COILED COIL (POTENTIAL).
FT DOMAIN 919 1607 COILED COIL (POTENTIAL).
FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
FT DOMAIN 1786 1817 COILED COIL (POTENTIAL).
SQ SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0FCB0 CRC64;
Query Match 4.58; Score 378; DB 1; Length 1818;
Best Local Similarity 19.68; Pred. No. 1.5e-05;
Matches 334; Conservative 308; Mismatches 595; Indels 470; Gaps 80;
QY 115 DNSSDSDAKS---YADLKHVRVLLTIKELKYPQ-----LFDLTNMLTLCDN 160
DB 32 ESAANDAQVVELQTQLLAEIKNLENEIKALKAQESQPDPHNNARQSLASLNRLVNE 91
QY 161 IHGPK-----YLIDGYERINELLYKLNPFYDLR-----AKLNDVCANDYQCIPFNLI 209
DB 92 YNNEFQKNYMWDRVAELNN---KAREFKDELKRLQENAAFNRYAN-WADFQSNYQL 147
QY 210 RANELDVLKLVGYRFPDNLKDNVGMEDYIKNNKTININE-----LIEESKKT 262
DB 148 KLDQFQAL-----IDQNOTIKQLNEQIAANQGLIDQNVQR 183
QY 263 IDKNKNAATKEBEKKLYQAQYDLISYKQLEEAHNLISVLEKRIDTLKKNENIKELDKI 322
DB 184 LQNHSLDQOERDALLYVDH---LYN-ELYELN---QKRLVGIEYEATQDLYSAD 234
QY 323 NEIKNPPANGNTPNTLLDNKKKIEBEKEIKEIAKTIFNIDSLFTDPLEYLYREK 382
DB 235 AELQNVYETIAQNANE---QKQCDAYWAQKQVEQIQ----- 270
QY 383 KNKIDISAKVETKSTENPNCVTPYPLSYND-----INNALNELNFGDLINPFYD-TK 437
DB 271 -----TTKQELVDEE-----STLKVRNLNDADFINSRLAELDDLTSKINERDFVSK 316
QY 438 EPSKNIVTDNRKRFINEIEKIEKIKETSDSKSYED-RKSLNDITK-EYEKLL--- 492
DB 317 EQADVV-----KASLANLTK-----EKERLSAEKDSFERLNTALNDINRWQENALFAK 366
QY 493 ---NEIYDSKFNNNIDLTFEE---KMMGKRY-SYKVEKLTHTHTTFASYENSKHNLEK --- 542

DB 367 HLEQOQVEFERKQOESLLKLETEHKQKRIGEPKIESEAKSEALLIQE---RELLERRE 424
QY 543 ----LTKA-LKYMEDYSLRNIVVEKE---LKYKNLI---SKIBENEIETLVENIKKDEEQ 591
DB 425 IDLLTQASLEYEQORRTNQVLKKEKROVQOQHFNQVHAKKKLDOKRHYLAEQRIDEEQ 484
QY 592 LFEKKITKDNKPKDEKILEVSDIVKVOQVLLMKNIDELKKTQILKLVNKLKHNHVPN 651
DB 485 IFK---LKEKIATERRELEKLYLVKKQK-----DQKENDLLIFEKQLR---QYQA 529
QY 652 SYKOENKQEPYLLIVLKEIKDKLVMPKVESLINEEKNIKITEG---SDNSESTEGEI 709
DB 530 DFENEIEEKQNELFASOKLSQKSFQTLKNKEAELNQAKIAEDWAHLKQNHHDLEI 589
QY 710 TQATTTPGOAGSALLEGDSVQAQO-----EOKQAQPPVPVPPEAKAVPT 757
DB 590 FLEGEFNHLOQEKHL---LEARTQFDNRVLSLSARFKQKAE-----LVKQKQSLQ 639
QY 758 PPAPVNNKTENV-----SKLDYLEKLYEFLNTSYICHKYILVSHSTMN---EKILQYKIT 810
DB 640 LTAAFNKEQEAVERDMDRLANLEKQEMLDGK---VHQF---DENSLSKLAERELAI 694
QY 811 KEESKLSKCDPLDLLFNINIPVMYSMFDLSNLSQLPMEI-YEKEMVNCNLYKLKDN 869
DB 695 KFEKELEAAQKQSLDN---NNAGLKLQDLKLSLTERLELEASKERILDFYD-ESS 751
QY 870 DKIKNLEEAKKYSTVSTKTLSSSMQPLSTPQDKPEVSANDDTSHST---NLANSKLKF 926
DB 752 RIADYESDLOARLAEVKTL-----KNOQETAASKERELKVALEKLNQAKAF 800
QY 927 -----ENILSLGKNKNTYQELIGOKSSSENFYKILKSDTFYNESFTNFVSKADDINSL 981
DB 801 LQIRKQOLLEIASVK---QOLAQKA-----NLLKNQQAELDKQ 835
QY 982 NDESKRKKLEEDINK---LKTTLQSLFDLYNLYKLKLER---LPDKKKT---VGKYMQ 1031
DB 836 TEELEAAFLQDQTKLEKALH---SVKSKQEL-LERERSFLLQKQREFAHVAGFKRQ 891
QY 1032 IKKTLLEKQLESKLSNPNPKHVLFQSV---FFNKKKEAEIAETENTLTKLLKHY 1088
DB 892 VHFKTQMQRL-SEFNQOQSQEQIKRETELKIAFADUKLDYQLFELQKNQBFQIQEKHK 950
QY 1089 K-GLVYKYNNGESSPLKTLSEESIQTEDNYASLENFVLSKL-----EGKLDN----- 1135
DB 951 ELELLAQKAE-----LQOELEOKATASQDQDTVQAKLDLARQOHELELRQAFNAQ 1004
QY 1136 ---LNLEKKKLSYSSGLHHLIAELKE-----VTKNNYTGNSPSENNTD---VNNALESY 1185
DB 1005 SLSLNKQREQLTNQVKVLH---GELKKRHEKLTLDRLAELAEKQDKHKKDAEIN---QRF 1058
QY 1186 KFLPEGTQV-----ATVVSSESGDITLQSQPKPKPASTHVGA 1223
DB 1059 KQFENEVADFQAKKRELQELNQTIRNLNQSASLLKKRNQTLTDFALLRK-----VQ 1111
QY 1224 SNTITTSQNVDDVDVVIPIFGESBEDYDQGVVTEAVTPSVIDNLSKINENEYEV 1283
DB 1112 HNTQTRVQLNTQIKEFL-----EKNQFQASDEAALQKALLIKRLRSFASKLQOREA 1166
QY 1284 LYLPLAGVYRSKLQLENNVMTFNVVNVDILNSRFNKRNF----- 1325
DB 1167 LAIQKLEFDRDEQKSEIN-----NAKLOEQFKLEKQNFDEAKOKLQIEFKDQCORL 1220
QY 1326 ---KNVLESDLIPYKDLTSSNYVVKD-----PYKFLN----- 1354
DB 1221 DVERLLKQVLQKLVNLSKSVLTITKNRADLSQQOQLQHYANLLEKLEKQAKRALDKKH 1280
QY 1355 -----KEKRDKFLSSYNIKD-----SIDTDI 1376
DB 1281 RAIYKMAQFVSELROEKQKLLSAQKQVDDKSRLLQENQRLQNLSSFTKKKROSLEHDI 1340
QY 1377 NFANDVLGYKILSEKYKSDLSIKKYINDKOGNEKYLPPLNNIET-----LYKTV 1428

Db 1341 N-----KFDQRRKEAVSILNSHKK-LKQEGELQGILOKLSKKTOIEQFESKLYQO- 1392
Qy 1429 NDKID-----LRFVHLEAKVLNTYEKSNVEYKIKELNVLKTIQ---DKLADFKKNNN-- 1478
Db 1393 REKLDRORTLSKLHRELKAQNEATAHKNREVLEIENYKKELQRLITTEKSEFDNKKRL 1452
Qy 1479 ---FVGIADLSDYDHNHLLTKFLSTGMVFENLAK---TVLSNLLDNGNQLMNTSQHQC 1532
Db 1453 FEYFRKI-----RNETEKEAHIKTVLEETQKRHLVETEAVKLHLKQKSIISKGOE 1504
Qy 1533 VKQCPQNSGCFRHLID-BREECKCLLN 1558
Db 1505 LKEIKERSVDRISHTNKKORELNSLLH 1531

RESULT 27

RA50_METJA STANDARD; PRT: 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 AtFase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Scott J.L., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Utterback T.R., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mreII by unwinding
CC and/or repositioning DNA ends into the mreII active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mreII (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U67572; AAB99331.1; -
CC TIGR; MJ1322; -
DR InterPro; IPR003439; ABC transport.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR Pfam; PF02483; SMC_C; 1.
DR ProDom; PD000006; ABC_transport; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SEQUENCE 1005 AA; 119387 MW; 9BBB48173E78F3 CRC64;

Query Match 4.4%; Score 372; DB 1; Length 1005;
Best Local Similarity 20.9%; Pred. NO. 1.2e-05;
Matches 272; Conservative 204; Mismatches 364; Indels 464; Gaps 67;
Qy 363 FNIDSLFTDP-----LELEYLYREKNKIDISAKVETKESTEPNE--YPNGVTVPLSYND 415
Db 56 FNYDTIITTKGKSVVVDLDFEVNGNNYKI-----IREYDSGRGGAKLYKNGPYATTISA 110
Qy 416 INNALNEL-----NSFGDLINPFDYTKPEPSKNIYTDNRKKFIN-EIKEKIKIEKKIES 469
Db 111 VNKAVNEILGVDRNPLNSI-----YIKQ-----GEIAKFLSLKPEKLETVAKLGI 158
Qy 470 DKKSIEDRSKSLNDITKEYKLLNIEYDSKFNNDIDLNFEMKMGKRSY---KVEKLT 526
Db 159 DE--FEKCYQKMEIVKEVEKRLERI-EGELNYK---ENYEKELKNKSQLSEKKNKLM 212
Qy 527 HNTFASYENSKHNLEKLTALKYMEDYSLRNIVVEKELKYYKNLISKIENETIETLVENIK 586
Db 213 IN-----DKLNKIRKEFED-----TEKLFNEWE----- 235
Qy 587 KDEOLFEEKITKDENK-----PDEKILEVSDIVKVOVKVLLMNK-----ID 629
Db 236 -NKKLLYEKFINKLEERKRALEKQELKILEYDLNTVVEARETLNRHKDEYKYSKLV 294
Qy 630 ELKTLQILKNVELKHNHVPNSYKQENKQEPYIYLVLKKEIDKLKVFMPK----- 680
Db 295 EIRKTESRLR--ELKSHY---EDYLLKLTQK----LEIIKGDIEKLKEFINKSKYRDDIDN 345
Qy 681 VESLINEEKKNIKTEGQSDNSEPSTEGETIGQATTKPGQOAGSALLEGDSVQAAQEOKA 740
Db 346 LDTLLNKIKDEIE-----RVETIKDLLEBK-- 371
Qy 741 QPPVPVPVPEAKAQPPTPPAVNNKTEENVSKLDYLEKLYEFLNTSVYCHIKYILVSHSTM 800
Db 372 -----NLNEIEKIEKRYCECKEY--EKL-----ELE 401
Qy 801 EKILKOYKITP-----BESKLSSCDPLD-----LFNIQNNIPWMSFOSLNSLSQ 849
Db 402 EKAVERNLTLEYITLLQEKKSIEKNINDLETRINKLLEETKNIDI-----ESIENSLK- 455
Qy 850 LPWEIYEREMWC-NLYKLK-----DNDIKNLLEAKKVSSTSVKTLSSSSMOP 896
Db 456 ---ELEEEKVLENOKEKIELNKKLGEINSEIKRLKILDELKEVEGKC-----P 503
Qy 897 LSLTPQD---KPEVSANDTSHSTNLNLSKLFLFENILSLGKNKIYQELIGOKSENFEY 953
Db 504 LCKTIDENKMKELI---NQHTQLNNK-----YTEL----- 532
Qy 954 KILKSDTFYNESFTNFVKSADDINSUNDESKRKLLEEDINKLAKTLQSLDLYNKYL 1013
Db 533 -----EIN-----KKIREIKDIEKKEIDKEENLKTPL 565
Qy 1014 KLERLFDKKTKYVKMOIKTLTLKEQLESKLSLNNPKHVLQNFSPFNKKKEAIEAE 1073
Db 566 YLEK-----QSQIEELEKLNKYNKEQDEINK---ISNVI---NKPVDEI--- 607
Qy 1074 TENTLENTKILLKHYKLVKYNGESSPLKLSIESIOTEDNYASLENFKVLSKLEGLK 1133
Db 608 ----LEDIKSQLNKFNFQVLSAVSYLSNVDEEGIR--NRIKEIENI-VSGWNKEKCR 660
Qy 1134 DNLNL---EKKKLSYLSGLHHLIAELKEVTKNKNYTGNSPENNTDVNNALLESYKFLP 1190
Db 661 EELNKLREDEREINLKOKLNELKNKEKELIE-----IENRRSLK--FDKYKEYL- 708
Qy 1191 EGTDVATVVVSESGDTELSQSQPKKPASTHVGAESNTITTSQNVDDDEVDDVIIVPIFGESE 1250
Db 709 -----GLTEKLEE-----LKNIKDGL-----E 725
Qy 1251 EDYDOLGQVVTGEAVTPSVVIDNLSKIENEYEVLYKPLAGVYRSLKOLENNVTFN 1310
Db 726 EYINCNSKIL-----AIDNKKRYNKEDIEY-----LNNKILEVNEK 764
Qy 1311 VKDILNSRPNKRNPNVLESDLIPYKDLTSSNYYVKDPYKFLNKEKDKFLSSYNIKD 1370

Db 765 INDI-----ERISVIN-QKLDEINYN--EEHKKIRELY-----ENKROELDNVREQKT 811
QY 1371 SIDTDINFANDVLGYKILSEKYSDDLSDIKKYINDKGENEKYLPFLNNIETLYKTVND 1430
Db 812 EIETGI-----EYLKDVESLKARLKE-----MSNLE-----KEKE 842
QY 1431 KIDLFVHLEAKLVNYYEKSNSVVKIKELNYLKTIO-----DKLADFKNKNFVGADLS 1486
Db 843 KLTFVEYLDK--VRRIFGRNGFOAYLRE-KYVPLIQKYLNEAFSEFDLPYSFV---ELT 896
QY 1487 TDYN-----HNNLIT-KPLSTGWMFENLA-----KTVLSNLLDGNL-----QGMNLIS 1528
Db 897 KDFEVRHAPNGVLTIDNLGG---EQIYAVALSRLAIALNIGRVECIILDEPTVYLD 953
QY 1529 QHOCVK-----KQCPONSQCFRHLDERECKCLLNYKQEGD 1564
Db 954 ENRRAKLAIEFRKVKSPQMIITTHRELEVDVADVIINVRKDN 997
RESULT 28
AKA9_HUMAN STANDARD; PRT; 3911 AA.
AC Q99936; Q9U0Q4; Q9U0H3; Q9Y6T2; O14869; O43355; O94895; Q9Y6B8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PRKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE (hyperion protein) (Yotiao protein) (Centrosome- and golgi-localized
PRK-associated protein) (CG-NAP).
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=98151389; PubMed=9482789;
RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
RT "Yotiao, a novel protein of neuromuscular junction and brain that
interacts with specific splice variants of NMDA receptor subunit
NRI.1";
RL J. Neurosci. 18:2017-2027(1998).
[2]
RN SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=99219864; PubMed=10202149;
RA Witczak O., Skallhegg B.S., Kerker G., Bornens M., Tasken K.,
RA Jahnsen T., Oerstavik S.;
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
protein located in the centrosome, AKAP450";
RL EMBO J. 18:1858-1868(1999).
[3]
RN SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
anchors multiple signaling enzymes to centrosome and the golgi
apparatus";
RL J. Biol. Chem. 274:17267-17274(1999).
[4]
RN SEQUENCE FROM N.A. (ISOFORM 1).
RP Kemmer W.A., Deiss S., Schwarz U.;
RT "Cloning of Hyperion";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric parietal cell;
RX MEDLINE=99115654; PubMed=9915845;
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.;
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
associated with centrosomes";
RL J. Biol. Chem. 274:3055-3066(1999).
[6]
RN SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC TISSUE=Lymphoblast;
RA Hinds K., Sutterer C., Becker M., Hawkins M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[7]
RN SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RC TISSUE=Lung;
RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
RT "AKAP350: A multiply spliced family of proteins with centrosomal
association";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[8]
RN SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro";
RL DNA Res. 5:277-286(1998).
[9]
RN SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
CC -1- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
(PKN). PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)
AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
CC -1- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND
CYTOPLASMIC IN PARIETAL CELLS.
CC -1- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3/CG-
NAP, 4/YOTIAO, 5 AND 6/AKAP350; ARE PRODUCED BY ALTERNATIVE
SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
CC -1- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC -1- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
FRAMESHIFTS IN POSITIONS 3782 AND 3811.
CC -1- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

CC EMBL; AJ131693; CAB40713.1; -
DR EMBL; AB019691; BAA78718.1; -
DR EMBL; AJ010770; CAA09361.1; -
DR EMBL; AF026245; AAB86384.1; -
DR EMBL; AF083037; AAD22767.1; -
DR EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR EMBL; AF091711; AAD39719.1; -
DR EMBL; AB018346; BAA34523.1; -
DR EMBL; AC000066; AAC60380.1; ALT_FRAME.

DR Genew; HGNC:379; AKAP9.
KW Coiled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-R11 SUBUNIT BINDING DOMAIN.
FT DOMAIN 164 314 COILED COIL (POTENTIAL).
FT DOMAIN 944 1022 COILED COIL (POTENTIAL).
FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).
FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).
FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).
FT DOMAIN 3726 3730 POLY-LEU.
FT DOMAIN 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT DOMAIN 1846 2772 GLU-RICH.
FT VARSPLIC 17 28 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPLIC 1637 1642 QLQBEI -> LATRD (IN ISOFORM 4).
FT VARSPLIC 1643 3911 MISSING (IN ISOFORM 4).
FT VARSPLIC 2175 2182 MISSING (IN ISOFORM 3).
FT VARSPLIC 2175 2183 SADTFQVKE -> Q (IN ISOFORM 3).
FT VARSPLIC 2895 2907 VGFYVMCFSTLC -> GSSIPELAHSDYOTREICSS (IN ISOFORM 2, ISOFORM 3 AND ISOFORM 6).
FT VARSPLIC 2895 2948 MISSING (IN ISOFORM 5).
FT VARSPLIC 3901 3911 STTQFHAGMRR -> ALSLTTSQWHSARPTAPLFEILSH (IN ISOFORM 6).
FT VARSPLIC 1347 1347 SLG (IN ISOFORM 6).
FT VARSPLIC 1347 1347 K -> QK.
FT VARSPLIC 76 76 /FTID=VAR_010926.
FT VARSPLIC 475 475 E -> Q (IN REF. 3).
FT VARSPLIC 554 554 M -> I (IN REF. 3).
FT VARSPLIC 638 638 E -> G (IN REF. 3).
FT VARSPLIC 663 663 R -> S (IN REF. 3).
FT VARSPLIC 913 913 N -> S (IN REF. 3).
FT VARSPLIC 956 956 H -> N (IN REF. 3).
FT VARSPLIC 980 982 K -> N (IN REF. 3).
FT VARSPLIC 997 997 QKH -> PKP (IN REF. 1 AND 2).
FT VARSPLIC 1001 1001 Q -> P (IN REF. 1 AND 2).
FT VARSPLIC 1020 1020 Q -> P (IN REF. 1 AND 2).
FT VARSPLIC 1028 1028 N -> D (IN REF. 3).
FT VARSPLIC 1626 1626 V -> E (IN REF. 3).
FT VARSPLIC 1703 1703 R -> P (IN REF. 1 AND 2).
FT VARSPLIC 1707 1707 V -> T (IN REF. 3).
FT VARSPLIC 1802 1803 V -> G (IN REF. 3).
FT VARSPLIC 1843 1843 MISSING (IN REF. 5).
FT VARSPLIC 1956 1956 A -> P (IN REF. 3).
FT VARSPLIC 2027 2027 I -> V (IN REF. 3).
FT VARSPLIC 2157 2158 V -> D (IN REF. 5).
FT VARSPLIC 2169 2169 EI -> HE (IN REF. 7).
FT VARSPLIC 2314 2314 E -> V (IN REF. 3).
FT VARSPLIC 2851 2851 L -> R (IN REF. 3).
FT VARSPLIC 2957 2957 I -> N (IN REF. 8).
FT VARSPLIC 2983 2983 E -> D (IN REF. 3).
FT VARSPLIC 3087 3087 P -> S (IN REF. 3).
FT VARSPLIC 3218 3218 Q -> H (IN REF. 3).
FT VARSPLIC 3307 3309 O -> H (IN REF. 3).
FT VARSPLIC 3751 3751 ESE -> QSQ (IN REF. 3).
FT VARSPLIC 3833 3833 P -> A (IN REF. 3).
FT VARSPLIC 3911 AA; 453664 MW; 3FBICB1C819B47AA CRC64; T -> S (IN REF. 3).
Query Match 4.3%; Score 365; DB 1; Length 3911;
Best local Similarity 18.3%; Pred. No. 8.9e-05;
Matches 285; Conservative 302; Mismatches 560; Indels 412; Gaps 69;
QY 173 E1NELLYKLNLFYF-----DLERAKLNDVCANDYQIQIPENLAKIRANELDVLKLVFGYRK 226
DB 242 ELTEQSQKLIQIQLOQASETLRNSTHSTADLLQAKQOQIILTHQOQLEQDHLLEDYQK 301

QY 227 PLDNKIDNVGKMDYIK-----KNKKTIIENINELIEESKTKIDKNKNATKEEKKKLYQQA 282
DB 302 KQEDFTMOISFLQEKIKVYVEMQDKKVENSKKEEOKEKTEIEELNTKIIEEKKTELEK 361
QY 283 YDLSIYNQKLEEAHNLISVLEKRIDTLK-----KNEIKELLDKINEIKNPP 330
DB 362 DKLTATDALLGELQIQVKNQOEIKNMKLELTSNOKERQSEEEIKQMGTVVEELQK--- 418
QY 331 ANSGNTPTLLDKNKK-----TEEHEKEIKEAKTIKFNIDSIFTDPL-ELEYLRE 381
DB 419 -----RNKDSQFETDIIVORMEQTORLEQLRAELDEMGQOIVQMQLR 466
QY 382 KKNKIDISAKVETKESTEPENYVTPYLSYNDI-----NNALNELN-SFGDIL 430
DB 467 QH-----MAQMEEMKTRHKGEMENALR---SYSNITVNEQIKLMNVALNELNIKLODTN 518
QY 431 NPFDYTKPSKNIYTDN-ERKAFINEIKIEKKIESDKKSYEDRSKSLNDITKE-- 487
DB 519 SOKELKEELGLILEKCALQLEDLVEELSFSRQIQARQRTIAEQESKLEAHKSL 578
QY 488 -YEKLLNLEYDS-----KFNNDIDLTFE---KMMGRYSYKVEKLTPHHTFASYNS 536
DB 579 TVEDLKAELVSASESKKELELKAHEAVNYKIKLEMLEKNAVLDRMA-----ESQ 630
QY 537 KHNLEKTKAL-----KYMEDYSLRNIYVEKELK-----YKKNLSKIEIENEIETLV 582
DB 631 EAELERLRTQLLSHEEELSCLKDELTEHRINIEKLDNGLHYKQKQDGLQNMESQKI 690
QY 583 ENIKDEQOLFKE-----KITK-----DENRDEKILEVSDIVK-----VQVOK 621
DB 691 ETMQFEKONLITKQNLILETSKLDKLOOSLVNSKSEMTLQINELQKEIETLRQEK 750
QY 622 VLLMKNKIDELK-KTOLIKNLVELKHN-----IHPNSYKQENKQEPYIIVLKKED 672
DB 751 GTLEQEVQELQKTELLEKQMEKENDLQEKFAQLAENSLKDEK-----KTLED 801
QY 673 KLVFMPKVESLINEKKNIKTEGSDNSEPT--EGEITQATTKPQQAGSAGLEGS 729
DB 802 MLKIHTP-----VSQERLIFLDSIKSKSDSWEKEIEIIEENEDLKQCC----- 848
QY 730 VQAAQAEQKQ-----AOPPPVPVPPEAKQ-----VPTPPAPVNNKTYNSKLDY--- 774
DB 849 IOLNEEIEKQRTTFAEKNFVNYQELQEEYACLLKVKVDDLEDSEKNOE---LEYKSK 904
QY 775 LEKLYEFLNTSYICHKYLIVSHSTMNE-----KILQYKITKEE 814
DB 905 LKALNEELHLQRINTTVKMKSSVFEDDKTFVAETLEMGEVVEKDDTTELMEXLEVTKREK 964
QY 815 SKLSS--CDPLDLFNQNNIPVMSFDSLNSLSOLFMEIYKEMVCLYKLD--- 868
DB 965 LELSQRSLDLSEOLKQKHGEISFLNEEVKSKOEQVSLRCRELEIINHNRRAENVQSC 1024
QY 869 NDKIKNLE---BAKKYSTSVKTLSSSSMOPLSITPODKPEVSANDDTSHSTLNLSL 923
DB 1025 DTQVSSLLDGVVMTSRGAEGSVKVNKSFGESKIWVEDK--VSPENMTVGEESKQEQ 1082
QY 924 KLFENILSLGKNKIYQELIGOKSSNFYKILKQSDTYPNES-----FTNP 970
DB 1083 -ILDHLPSVTKESS---LRATQPSN--DKLQKELNVLKSEQNDLRLQWEAQICLSLV 1135
QY 971 VSKAKDDINSINDESKRKKL-----EEDINKLAKTLQLSFDLYKYLKLERLF 1019
DB 1136 YSTHVDQVREYMEKDKALCSLKEELIFAQEKIEKQKIHOLEQ-----TMKTQETG 1190
QY 1020 DKKK---TVGKYKMOI-KKLTLLKEQESKLSLNNP-----KHVLQNFVSFFNKKK 1067
DB 1191 DGKPLHLLIGLKQAVSECSYFIQTLCSVLGEYTPALKCEVNAEDKENSODYISENE 1250
QY 1068 EAEIA---ETENTLNTKILL---KHYKGLV-----KYNGESSPLK-TLSESTQ 1111
DB 1251 DPQLQRYREVQDFQENMHMTLLNKVTEEYNKLLVLQTLRSLKIVGQQTGCMKLEFGEENLP 1310
QY 1112 TED-----NYASLENPKVLSLEGKGLKLDNLNLEKKKLSYLSGLHLHIAELKEVKN 1163


```
Db 1311 KEETFLSIHSONTLEIDVNHK--SKLSSLDQLEKTKLEQVOELESLSLQQQLKE 1368
Qy 1164 KNYTSGNSPENTDNNNALESYKKFLPEGTDVATVVSSESGDTLQSQPKKPASTHVGA 1223
Db 1369 -----TEQNYAE--IHCLKRL-----QAVSES-----TVPSLPVDSVWITE 1405
Qy 1224 SNTITP-----SONVDEDDVIVIPFGSEEDYDGLQGVVTGEAVTSPVIDNLSK 1276
Db 1406 SDAQRTMYPGSCVKKNDIGTIE--FSGEFGVKEE-----TNIVKL 1443
Qy 1277 TENEYEVLYLPLAGVYKSLKOLENNVMTFNVNWKDILNSRFNKRNFKNVLESDLIPY 1336
Db 1444 LEKOYQ-----EQLSEVAKVIVS-----MSTAFQAQOTELSRISGG----- 1479
Qy 1337 KDLTSSVYVKKPYKFLNKRDKFLSSYNYTKDSIDTIDINFANDVLGY--YKILSEYK 1394
Db 1480 KENTASS---KQAHAVCOEQ-----HYF-----NEMKLSQDQIGFQFETVDVAFK 1523
Qy 1395 SLDLSIKKIYNDKQENKYLPLFLNN-----IETLYKTVNDKIDLVFVHLEAKV 1443
Db 1524 BEFKPLSKEL-----GEHGKEILLNSDPHDIPESKDCVLTITISEEMFSKDKTFIVR----- 1574
Qy 1444 LNYTEKSNVVEKIKELNYLXTI---QDKLADFKNNNFVGTADLSTDYNNHNLTKFL 1499
Db 1575 -----QSHDEISVSSMDASRLMLNEQLEDMRO-----ELVRQYQEHQQAPELL 1620

RESULT 29
DMD_CHICK
ID DMD_CHICK STANDARD; PRT; 3660 AA.
AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dystrophin.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID-9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE-89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE PLASMA MEMBRANE.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN, ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC
CC This SWISS-prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```
CC -----X13369; CAA31746.1; -
DR PIR: S02041; S02041.
DR HSSP: P46939; 18HD.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR002349; WW.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR InterPro; IPR000433; Znf_Z2.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00397; WW; 1.
DR Pfam; PF00435; spectrin; 22.
DR Pfam; PF00569; Z2; 1.
DR PRINTS; PR00403; WWDOMAIN.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEG; 21.
DR SMART; SM00456; WW; 1.
DR SMART; SM00291; Znf_Z2; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
DR PROSITE; PS01357; ZF_Z2_1; 1.
DR PROSITE; PS0135; ZF_Z2_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; Zinc-finger.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 19 123 CH 1.
FT DOMAIN 138 241 CH 2.
FT REPEAT 341 449 SPECTRIN 1.
FT REPEAT 450 558 SPECTRIN 2.
FT REPEAT 561 669 SPECTRIN 3.
FT REPEAT 721 830 SPECTRIN 4.
FT REPEAT 832 936 SPECTRIN 5.
FT REPEAT 945 1047 SPECTRIN 6.
FT REPEAT 1050 1156 SPECTRIN 7.
FT REPEAT 1159 1265 SPECTRIN 8.
FT REPEAT 1268 1369 SPECTRIN 9.
FT REPEAT 1470 1570 SPECTRIN 10.
FT REPEAT 1573 1678 SPECTRIN 11.
FT REPEAT 1681 1782 SPECTRIN 12.
FT REPEAT 1879 1981 SPECTRIN 13.
FT REPEAT 2013 2103 SPECTRIN 14.
FT REPEAT 2106 2211 SPECTRIN 15.
FT REPEAT 2214 2321 SPECTRIN 16.
FT REPEAT 2472 2574 SPECTRIN 17.
FT REPEAT 2577 2683 SPECTRIN 18.
FT REPEAT 2686 2799 SPECTRIN 19.
FT REPEAT 2802 2904 SPECTRIN 20.
FT REPEAT 2906 2928 SPECTRIN 21.
FT REPEAT 2931 3037 SPECTRIN 22.
FT DOMAIN 3052 3085 WW.
FT ZN_FING 3304 3351 Z2-TYPE.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

Query Match 4.3%; Score 361; DB 1; Length 3660;
Best Local Similarity 18.8%; Pred. NO. 0.00011;
Matches 343; Conservative 297; Mismatches 592; Indels 588; Gaps 82;

Qy 140 KELKYPQLFDLTNHLMT-----LCDNTHG-FKYLIDGYEEN 175
Db 1212 KEVKVKLTDSYNNPIAKAPPAANEALKKELDLVLTYSQRLGSKCKTL-----EEVW 1267
Qy 176 ELLYKLVNFYDLRLAKRNDVCANDYQIPFNLIKIRANE-----LDVKKLVFG 223
Db 1268 ACWHELLSYLDAENKWLNEV-----ELKATENIQGAEISELSLERLM-- 1315
Qy 224 YRKPLDNIKDNVGNMEDIYKKNKKTENINELIE-----SKKTIDKNNA 269
```



```
Db 1316 -RHPEDN-RNQIRELAQTLDGGILDELINLEKLENTWEELOQAVRRQKSLEQSIQS 1373
QY 270 TREEEKKLYQAQYDLSYNNKQL-----EAAHNLISVL---EKRIDTLKK- 311
Db 1374 AOETD-KTLRLTOESLAADKQITAYTADRVDAQAQVPOBAQKIQSLETSHETSLEEMKKR 1432
QY 312 ---NENIKELLKINIKNPPANSQNTPTLLDKNKKTEEHEKEIKEIAKTIKFNIDSL 368
Db 1433 NRKESAKRVLSQIDVAOK--KLQDYSMKFRLEFQKPAFQORLOECKRILDEVKLOVPKL 1490
QY 369 FTDPLE-----LEYLREKKNKIDISAKVET-----KESTE-PNEYPNGVT-Y 409
Db 1491 ETKSVQEVQVSHLDHCKMLKLSLVSEVKSEVETVIKTGRQIVQKQOQTEPNKPELDERLTAL 1550
QY 410 PLSYNDINNAL-----NELNSFGDLINPED----- 434
Db 1551 KLOYNELGAKVTEKKQELKEKLUKLSRKLRKEINLSFEWLAATDVLTAKRSVQOGMPSNLD 1610
QY 435 ---YTKPSKNIYTDNERKKFTINEIKIEKKKIESDKSYEDRSKSLND-----ITK 486
Db 1611 AETAWGKATRKETKRQVQLKNICDILGENLKVTLKGKES--LVEDKLSLLNSNWTAVTS 1667
QY 487 EYEKLLN-----EYDSFKNNNIDLTNFKMMGKRYKYKVEKLTHHNTFASYENSK 537
Db 1668 RAEWLNLLMEXQKHMEAFDQVAN--VTT-----WIYRAETILLDESQKRPQKE 1716
QY 538 HNLEKLTALKVMEYD--SLRNIVVE----- 561
Db 1717 ETLKRLKAEINMHPKVDVSRQAVDLMNTRGDHCRKVTEPKLSELNHRFAISQRIKSG 1776
QY 562 -----KELKY-----KNLISKIENIEIIVENIKKDEQLPEKKTIKDKENPDKILEV 611
Db 1777 KPPIPLKELEQDFDIQKLEPLEVEIQGV-NLKEED--FNKQMSDEESTVKELLQR 1832
QY 612 SDIVKVOVKVLLMKNKIDELKKTOLLKN--VELKINI--HVPNSYKQENKQEPYLYIVL 667
Db 1833 GPTLQKRI-----TDERKREEIKIKOQLQTKHNAKLRSQRRKKALEISHOWYOY 1884
QY 668 KKEIDKLVKMPKVESLI-----NEBKNIKTEGQSDNSEPSTEGEITQAVTKPGQOAG 722
Db 1885 KRQADDLMTWLDIEKKLASLPDHKDEQKLEIGGELEKKEDLNNAVQAEERLSKDGAA 1944
QY 723 SALEGDSVQAQAQ-----EQQAQPPVPVPVPEAKAQVTPPAPVNNKNTENVSKLDYLEKL 778
Db 1945 KAVEPTLVQLSKRWDRFESKFAQ-----FRLNVAQIQ 1977
QY 779 YEFLNTSYCHKIYLS-----HSTMNEKILKQYKITEESKLS-----CDPL----- 823
Db 1978 TVLEDTFVMTESMTVETTYVPSTYLAETLQLLQALSEVEERLNSPVLOAKDCEDLLKQE 2037
QY 824 -----DLLFNIONNIPWYIS-MFDSLNNLSLSOLPMEIYEK-----EMVCNLYKIK 867
Db 2038 ECLKNIKDCGLRGHIDIIHSKKTALQASATPRETANTQDKLTQLNSQWEKVNKYRDR 2097
QY 868 DN--DKIKNLEPAKKVSTSVKTL-----SSSMQPLSLTPQDPKPEVSANDTSHSTNLNS 922
Db 2098 QARFDRSK--EKWRLFHCMSKSNWLTETETEKLS-----RAQIEAG-DVGHV----- 2142
QY 923 LKLFENILSGKNKIYQEL--IGQ-----KSNSEFYKILKDSDTFYNESFTNFVSK 974
Db 2143 -----KTKQFLQELQDQIGRQQTAVTKTANVTGEEIIEQS-----SAADANVLEKQ 2187
QY 975 ADDINSINDB-----SKKKLE-----EDINKLKKTLQLS----- 1004
Db 2188 LGNLTNRWQIECBQVEKKRIEENKNIISFQEDLNKLILWLEETENVIAIPLPCGNE 2247
QY 1005 --FDLYNKYKLERLDFKK-----KTVGKYKMQIKKLTLLKQESKLSNLSLN 1051
Db 2248 QLRDCGLGKVRVEELLPHKILKRLNETGGTTLGASLNPER-----KHKLESTLKEAS- 2302
QY 1052 PKHVLQNSVFVNKKEAEI-----AETENTLNTKILKHYKGLVKKYVNGESSP-LKTL 1105
```

```
Db 2303 -RLLKVSRLPEKQKEIILLKDFIELNQOINQLTWTIPVKNQLELYNQVQPGAFDI 2361
QY 1106 SESIQOTEDNYASLENFKVLSKLEG---KLKDNLNEKKKLSYLS--GLHHLIAELK 1158
Db 2362 KETEAQVQAKQPNVE--EVLSK--GCHLYKEKPAHPVKKKLEDNADWKAINHLLQLK 2417
QY 1159 E-----VIKKNYITGNSPSENNTDNNALLESYKFKFL 1189
Db 2418 EKPTFGEPALTPSGVLTSGTVAVDTPQARVTKETTSFTPTMPSSVLEEV-PALADENKAW 2476
QY 1190 PGTDVAT-----VVSESGDTLEQSQPKKPASTHVGAESNT 1226
Db 2477 AELTDLWSLRDREIKAQRVTVGDLDDINDMIKQKANMODLEQRRPQL-----DEL 2527
QY 1227 ITTSQNVDDDEVD---VLIIVPIFGSEEDYDDL-----GOVVTGEA 1264
Db 2528 ITAAQNLKNTSNOEARTIITDRIEKIQSOWDDVHGYLQNRROQLHEMOKDSTQWLEAQ 2587
QY 1265 VTPSVIDNLSIKIENEYEVLY-----LKPAGVYVRSLLKQLEN----- 1302
Db 2588 EAEQVLEQAKAKLESWKESYTVTEALKQNSLQKQSKETIRQWQMNIEGVNDVALKPVRD 2647
QY 1303 -----NVMTFNVNVR-DILNSRFNKRENFKNVLESOLIPYKDLTSSNVVVKDPYKF 1352
Db 2648 YSADTRKRVLMETDINATWATINKRVSERE---AALESALLMLQEF-----Y 2692
QY 1353 LNKERDKFELSSYNIKDSIDTIDINFANDVLYGYKILSEKYSKSDLSIKYIINDKQGENE 1412
Db 2693 LDLEKFLAWLT-----EAETTANVLQDATHKEKTLDPQM--VRELMKQWDLQAEID 2743
QY 1413 KYLPFLNNI-----ETLYKTVNDKIDLFVHLEAKVLYNYEYKSNVEVKIKELN---YLK- 1464
Db 2744 AHTDIPHNLDENGOKILRSLEGSSEDAVLLQRRLDNNFRWS-----ELRKSLNIRSHLEA 2799
QY 1465 -----TTODKLA-----DFKKNFNVGIADLSTDYNNHNLITKF---LST-GMV 1504
Db 2800 STDQWRKRLHLSQELLANWLQKDEKLQQAQPIG-GDPIPTVOKONDVHRTFKRELKTEPV 2858
QY 1505 FENLAKTVLSNLIDGNLQGM 1524
Db 2859 INNALETVRLELADQPVGL 2878

RESULT 30
ID TPR_HUMAN STANDARD; PRT: 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=14371155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Sapphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
```


QY 1377 NPAVDVGLYKILSK-----YKSDLSIKKYINDKGENEYLPFLNN 1420
Db 1352 E-----EYRKLSEVHTKRIQOLTTEEIGRLKAEIARSASLTNNQIOLSKEDLNK 1405
QY 1421 IETLYKTVNDKIDLVHLEAKVLYNT-----YKSNVEVKIKELNYLKTIOQKLD 1472
Db 1406 VTEKETIQKIDAKIIDIEQVKVITQVKKIGRRYKTOYELKQAQDKVMTSAQSGD 1465
QY 1473 FKNNNFV-GIADLSTVDYNNHLLTKFLSTGMVFENLAKTV-----LSNLLDGNLQGLM 1525
Db 1466 HQEHVSQVQEMQELKTNLQATKSKLSQV--ENLQTLSEKETEARNLQEQIVQLOS 1523
QY 1526 NISQ-HQCVKKQCPNSGFRHLDRER--CKLLNYKQE-----GDK-CVENPNTONE 1576
Db 1524 ELSRLRQDLQDRTTQEQRLRQOITKEKTRKAIKAAKSTIAHLAGVKDQLTKEEELQ 1583
QY 1577 NNGGCD-----ADAKCTEDSGNGK 1597
Db 1584 RNLGDDQKQDELVDRLTALKSQYEGR 1609

RESULT 31

CENP_HUMAN STANDARD; PRT: 3210 AA.
AC P49454; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH antigen).
GN CENPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis.";
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B., Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal domain sufficient for nuclear localization.";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization, centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN

CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH RETINOBLASTOMA PROTEIN (P130), CENP-E AND BUBR1.
-!- SUBUNIT: HOMO- OR HETERODIMER.
-!- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS), REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
-!- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
-!- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U19769; AAA82889.1; -;
DR EMBL; U30872; AAA82935.1; -;
DR EMBL; U25725; AAA86889.1; -;
DR Genbank; HGNC:1857; CENPF.
DR MIM; 600236; -;
DR InterPro: IPR001230; Prenyl site.
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil; Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1328
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2568
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT DOMAIN 3015 3032
FT CONFLICT 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2243
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2561
SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
Query Match 4.2%; Score 355.5; DB 1; Length 3210;
Best Local Similarity 18.8%; Pred. No. 0.00015;
Matches 349; Conservative 305; Mismatches 631; Indels 569; Gaps 79;
QY 125 YADLKHRR--VRNVLTKELKYPQLFDLTNHLMTLCDNIHGFYKLDG---YEEINELLY 179
Db 709 FSDQKHQKEIENWCKTSQLT-GQVEDL-EHKQLLSN-----EIMDKRCYQDLHAHYE 761
QY 180 KLNIFYDLLRAKLNDVCAAND-----YCQIPFNLIKIRANELDVLKLVFGYGRKPLDNI 231
Db 762 SLR---DLLKSKDASLVTNEDHQRSLLAFDQQPAMHHSFAN-----IIGEQGSMPS 810
QY 232 KDNVGMEDYIKNNKTTENINELTEESKTTDKNNKATKEBEKKLYQAQYDLSIYNKQ 291
Db 811 RESECLADQSPKSAIQLQNRVDSLEFSLES-QKQMSDQKQCELVQIKGEIEENLMK 869
QY 292 LSEAH-NLISVLEKRTDITLKN-----ENIKELDKINEIKNPPPPANSNTENTLL 341
Db 870 AQMHQSFAETSQRISKLQEDTSAHQNVVAETLSALENKEKELQ-----LL 916
QY 342 DNKKIEEKEIKEITAKTIKENIDS-----LFTDPLELEYLYREKNNKIDISAKVETKE 396
Db 917 --NDKVETEQAIEQLKSNHLLDSLSKLQLSLTSLSE--KKEMSSIISLN-KREIEE 971
QY 397 STEPNEYPNGVITYPLSYNDINNALNELNSFGDLINFDYTKEPSKNIYTDNKKKFINI 456

RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Complete nucleotide sequence of full length cDNA for rat beta
 RL cardiac myosin heavy chain";
 RN Nucleic Acids Res. 17:7529-7530(1989).
 RP [2]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE=90133919; PubMed=2614840;
 RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
 RT Comparisons suggest a molecular basis for functional differences";
 RL J. Mol. Biol. 210:665-671(1989).
 RN [3]
 RP SEQUENCE OF 1524-1935 FROM N.A.
 RX MEDLINE=82220036; PubMed=7045682;
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
 RT "Molecular characterization of two myosin heavy chain genes expressed
 RT in the adult heart";
 RL Nature 297:659-664(1982).
 RN [4]
 RP SEQUENCE OF 1871-1935 FROM N.A.
 RC STRAIN=Wistar; TISSUE=Heart;
 RX MEDLINE=85179510; PubMed=6241892;
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac myosin heavy chain isozymic transitions during development
 RT and under pathological conditions are regulated at the level of mRNA
 RT availability";
 RL Eur. Heart J. 5:181-191(1984).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL; X15939; AAA34065.1; -;
 CC EMBL; J00752; AAA41654.1; -;
 CC EMBL; M32698; AAA41659.1; -;
 CC PIR; S06006; S06006.
 CC PIR; A02989; A02989.
 CC HSP; P08799; IAMD.
 CC InterPro; IPR000048; IQ_region.
 CC InterPro; IPR004009; Myosin_N.
 CC InterPro; IPR002928; Myosin_tail.
 CC InterPro; IPR001609; myosin_head.
 CC Pfam; PF00063; myosin_head; 1.
 CC Pfam; PF00612; IQ; 2.
 CC Pfam; PF01576; Myosin_tail; 1.
 CC Pfam; PF02736; Myosin_N; 1.
 CC PRINTS; PD00193; MYOSINHEAVY.
 CC ProDom; PD000355; MYOSINHEAVY.
 CC SMART; SM00015; IQ; 1.
 CC SMART; SM00242; MYSC; 1.
 CC PROSITE; PS50096; IQ; 1.

KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 780 MYOSIN HEAD-LIKE.
 FT DOMAIN 781 810 IQ.
 FT DOMAIN 840 1935 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP.
 FT DOMAIN 655 677 ACTIN-BINDING.
 FT DOMAIN 757 771 ACTIN-BINDING.
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 695 695 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 1529 1531 IRK -> VRR (IN REF. 3).
 FT CONFLICT 1731 1731 D -> H (IN REF. 3).
 FT CONFLICT 1784 1784 N -> K (IN REF. 3).
 FT CONFLICT 1851 1851 T -> N (IN REF. 3).
 FT CONFLICT 1858 1858 R -> K (IN REF. 3).
 FT SEQUENCE 1935 AA; 223082 MW; C8376C324A9BD82B CRC64;
SQ
 Query Match 4.2%; Score 351.5; DB 1; Length 1935;
 Best Local Similarity 18.9%; Pred. No. 0.00011;
 Matches 286; Conservative 265; Mismatches 589; Indels 377; Gaps 62;
 QY 28 LVKLEALEDAVLGYSLFQEKVMVNEGTSCTAVTSTPGSKGVSASGGSGSVASGGS 87
 DB 594 LOKNKDPLNETVV--GLYQKSSLLKLLSLNFANYAGADAPVDKKGK-----GK 637
 QY 88 VASGGSVASGGSVASGGS-----NSRRTNPSDNSSSDAKSYADLKHRYNVLITIKELK 143
 DB 638 AKKGSFQVTSALHRENKLNKMTNLRSTHP-----HFVR--CIIPNETK 679
 QY 144 YPOLFD--LTNHLMTLCDNIHGFYLDGYEINELLYK-----LN-----FYFD 186
 DB 680 SPGVMDNPLVMHQLRCNGVLEGIRICRKGFP--NRILYGFQRYRILNPAAPPEQGFID 737
 QY 187 LLRA--KLMDVCANDYQIQPFN-----KIRANELDVLK 218
 DB 738 SRKGAELKLSGLDIDHNQYKFGHTKVFYFAGLLGLLEEMRDERLSRIITRIQAQSRGLVS 797
 QY 219 KLVFVGKPLD-----NIKDNVG-----KMEDYIK--KNKKTIEINEL 255
 DB 798 RHEF--KLLERDLSLIITQWIRAFMGVKNPWNPKLYFKIRPLKASATEKEMANKKEE 855
 QY 256 IEESKTTIDKNKNATKEEKK--KLYQAQYDLSYKNQLEEAHNLISVLEKRTIDTLKKNE 313
 DB 856 FGRVDALEKSEARKELEKVMVSLQERNDLQL---QVQAEODNLADAEERCDQLKNNK 912
 QY 314 -----NIKELDKINEKNPPANSNGTNTNTLLDKNKKIEEHEKEIKEAKTIKFNIDSL 368
 DB 913 IQLEAKVKEMTERLEDEE-----EMNAELTAKKRKLEDECSSELKRDID-- 955
 QY 369 FTDPLELEYLREKNKNIDISAKVE--TKESTEPNYPNGVTYPLSYNDINNALNELNSFG 427
 DB 956 -----DLEITL-----AKVEKEKATE-----NKVKNLTEEMAGLD 986
 QY 428 DLINPFDTYTKPSKNYITDNERKKFNETKEIKIEKKIESDKKSYEDRSKSLNDITKE 487
 DB 987 EII--VKLTKE-----KKALQEAHQ--ALDDLOAEEDK-----VNTLTKA 1023
 QY 488 YEKLLNEIYDSKFNENNIDLTNFEKVMGKYSYKVE---KLTHHTNTFASYSKHNLEKLT 544
 DB 1024 KYKLEQQQVDD--LEGLSDQDKKVRMDLRAKRKLEGLDKLT--QESTMDLENDKQQLDERL 1080
 QY 545 KALKY-----MEDYSLRNITVVEKELKYKNLKIENIETLTVENTIKKDEQLEFEK 596
 DB 1081 KKKDFELNALNARIEDEQALGSQLKKLQELQARIEELEEELE--AERTARAKVEKLRS 1138
 QY 597 ITKDNKPKDEKILEVSDIVKQVQKVLNNK--IDELKKTQLTLKVKLKHNIHVNSYK 654
 DB 1139 LSRELEIEISERLEEAGGATSQIE-----MNKREAEFQMRDRDLEATLQHEA---TAA 1190
 QY 655 QENKQEPYLIIVLKKEIDKLKVFMPKVESLINEEKNKIKTEGSONSESTEGEITGQAT 714


```
Db 1191 ALRKHADSVAELEQIDNLO-----RVQKLEKESEFKLE--LDDVTSNMBEQIIKAKAN 1244
Qy .715 ----TKPGQAGSALGDSVOAQAOEKOKAOPVPVPVPEAKAQVTPPPAPVNNKTN 768
Db 1245 LKMCRTLEDQMNHRKAEETQSRVNDLTQRAKLQTEGELSQOLDEKEALISQLTR- 1303
Qy 769 VSKLDYLEKLYEFLNTSYCHKYIILVSHSTMNKILKOYKITKEESKSSCDPLDLFN 828
Db 1304 -GKLTYTQOEDL-----KQLEEVKAKNALAHALQSAHDCDLDLREQVE 1348
Qy 829 IQNNIPVMYSPDLSNLSOLFMEIYKEMVCNLYKKD-NDKIKNLEEAKKVSTSVK 887
Db 1349 ETEKAELQRLVSKANSEVAQMRKYETDAIQRTTELEEAKKKLAQLQDAEAEAVN 1408
Qy 888 TLSSSSMPLSLTPQDKPE---VSANDDTSHSTNLSNLSKLFENILSGKKNKNIQEILG 944
Db 1409 A-KCSSLEKTKHRLQNEIEDLMVDYERSNAAAALDKKORNFDKILVEWKQK--YEEQS 1465
Qy 945 QKSS-----ENFYKILKSDTFYNESFTNFVKSADDTINSLND----- 983
Db 1466 ELESSQKEARSLSLTFELFKLNAYEESLEHETFKREN-----KNLQEEISDLTEQLGSTG 1520
Qy 984 -----ESKRKLEEDINKLTKL-----QLSPDLNKKYKIKLER-LF 1019
Db 1521 KSIHELEKIRKOLEAEKLELQSALEAEASLEHEGKILRAQLEF---NQKAEIERKLA 1577
Qy 1020 DKKTGVGYKM-QIKKLTLLKEQESKLSLNNPKHVLQNFVFFNKKKKEAIEAETENTL 1078
Db 1578 EKDEMEQAKRHLRVDSLSLQTSDAETSRNEALRV-----KKMEGDLNEMETQL 1629
Qy 1079 ENTKILLKHGKLVKYYNGESSPLTLSEESIQTEDNYASLENFVKSLKLEGLKDNLNL 1138
Db 1630 SHANMAEAQKQV-----SLQSLKDTQIQDDAVRANDD-----LKENIAI 1673
Qy 1139 EKKKLSYLSGLHLLHIAELKEVKKNNKNTGNSPSNNTDV-----NNALSEYKFL-PE 1191
Db 1674 VERRNNLQAELEELRAVVEOTERSKLAEOELIETSERVOLLHQSNTSLNOKKKMDAD 1733
Qy 1192 GTDVATVYSESGDFLEQSPKPPASTHVGAESNTITTSQ-----NVDDEVDDV 1240
Db 1734 LSOLQTEVEEAQVCEAEEAKKAITDAAMWAELKKEQDTSALHRKKNWETIKDL 1793
Qy 1241 IIVPIGSESEYDDLQGV-VTGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLKQ 1299
Db 1794 -----QHRLEAEQIALKGKQLQKLEARVRELENELE-----AEQKR 1832
Qy 1300 LENNVMTNVNVKDLNLSRFNKNFNKLVLESDLIPYKDLTSSNVVVKDPYKFLNKEKD 1359
Db 1833 NAESVKGMRKSRRIKELTYQTEEDRKNLLR-----LQDLVDKQLQKVKAYKQAEAE 1887
Qy 1360 KF---LSSYNYIKDSID 1373
Db 1888 QANTLSKFRKQVHOLD 1904

RESULT 35
MYH7_HUMAN STANDARD; PRT: 1935 AA.
AC P12883; Q14904; Q16579;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MYHC-beta).
GN MYH7 OR MYHCB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91065634; PubMed=2249844;
RA Jaenike T., Diederich K.W., Haas W., Schleich J., Lichter P.,
RA Pfordt M., Bach A., Vosberg H.P.;
```

```
RT "The complete sequence of the human beta-myosin heavy chain gene and
a comparative analysis of its product.";
RL Genomics 8:194-206(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90301496; PubMed=2362820;
RA Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,
RA Anderson D.H., Lin L., Liew J.;
RT "Complete sequence and organization of the human cardiac beta-myosin
heavy chain gene.";
RL Nucleic Acids Res. 18:3647-3651(1990).
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1310-1935 FROM N.A.
RX MEDLINE=86176778; PubMed=2421254;
RA Saez L., Leinwand L.A.;
RT "Characterization of diverse forms of myosin heavy chain expressed in
adult human skeletal muscle.";
RL Nucleic Acids Res. 14:2951-2969(1986).
RN [6]
RP REVISIONS.
RA Leinwand L.A.;
RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 1410-1935 FROM N.A.
RX MEDLINE=88299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
beta-form myosin heavy chain complementary DNA clones. Regulation of
expression during development and pressure overload in human
atrium.";
RL J. Clin. Invest. 82:524-531(1988).
RN [8]
RP SEQUENCE OF 785-1935 FROM N.A.
RT TISSUE=Skeletal muscle;
RX MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
RA Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [9]
RP SEQUENCE OF 1393-1935 FROM N.A.
RX MEDLINE=87192738; PubMed=3032769;
RA Jandreski M.A., Liew C.-C.;
RT "Construction of a human ventricular cDNA library and
characterization of a beta myosin heavy chain cDNA clone.";
RL Hum. Genet. 76:47-53(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=96039076; PubMed=8533830;
RA Arai S., Matsuoka R., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
RA Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
RA Hosoda S., Momma K.;
RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
hypertrophic cardiomyopathy.";
RL Am. J. Med. Genet. 58:267-276(1995).
RN [11]
RP VARIANTS CMH1 GLU-256 AND ARG-741.
RX MEDLINE=93248216; PubMed=8483915;
RA Panapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;
RT "Missense mutations in the beta-myosin heavy-chain gene cause central
core disease in hypertrophic cardiomyopathy.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
```


Db	956	----	DLELTL-----	AKVEKKUATE-----	NKVKNLT	TEMAGLD	986	
Qy	428	DLIMPEDY	TKREPSKN	ITYDNERK	KFFINEIK	EIKIEKK	KKIESDKSYEDRSKSLNDITKE 487	
Db	987	EIIAKL--	TKE-----	KKALOE	AHQ--	ALDDQL	QAEDK-----VNTLTKA 1023	
Qy	488	YEKLLN	EYDSKEN	NNIDL	TNEK	MMGKRY	SKYVE---KLPHNTTFASYENSKINLEKLT 544	
Db	1024	KVLEQO	VD--	LEGSLE	QEKYVM	DLERAKR	KLEGDULKT--QESIMPLENDKQOLDERL 1080	
Qy	545	KALKY-----	----	MEDYS	LRNI	VEKEL	YKKNLYSKLISKELETELVENIKKDEQLPEKK 596	
Db	1081	KKDFEL	NALNARI	EBEQ	ALGSOL	QKKL	KELQARTEEELELES---ERTARAKVEKLRSD 1138	
Qy	597	ITKDEN	KPDKILE	VS	DI	VKVQ	VYLLMKN--IDELAKTQILIKNVELKKNIHVPNSYK 654	
Db	1139	LSRELE	ISERLE	EAGG	ATSVQIE	-----	MNKKREAEFQKMRRLDEATLQHEA---TAA 1190	
Qy	655	QENKQEP	YI	LVIL	KEID	KL	KVMPKVESLINNEKKNIKTGQSDNSPSTEGEITQAT 7144	
Db	1191	ALRKKH	ADSVAEL	GEQID	NLQ---	RVKOK	LEKSEFKLE--LDVTSNMNEQIIKAKAN 1244	
Qy	715	-----	TKPQQA	GS	AL	SGDS	VQAQAEQOQAPVPVPVPEAKAOVPTPPAPYNNKTEN 768	
Db	1245	LEKMCRT	FLEDMN	EHRSK	AETQ	RSVND	ITSORAKLOTENGELSQOLDEKALISQLTR- 1303	
Qy	769	VSKLDY	LEKLYE	FLNT	SYCH	KYIL	VSHSTMKEKILQYKITKBEESKLSLSCDPLDLLFN 828	
Db	1304	-GKLT	YTOQ	LEDL-----	----	KQLEEV	EYKARNALAHALOSARHCDLLRQYE 1348	
Qy	829	IQNNIP	VMYS	SMFDS	LANS	LSQ	FMELYEKEMVCNLYKUKD--NDKTKNLEBAKKVSTSVK 887	
Db	1349	EETEA	KAEAL	QVLRS	KANSE	VAQWRT	KYETDAIQRTEELEEAKKLAQRLQEAEEAVEAVN 1408	
Qy	888	TLSSSS	MOPLSL	TPQDK	PE---	VSAND	DFSHTNLSNLKFLNLTLSLGGKNKYIQEILG 944	
Db	1409	A-KCSS	LEKTKH	RQNE	IEDL	WV	VERSNAALADJKQRFNDKTLAEWKQ--YEESQS 1465	
Qy	945	QKSS-----	----	ENFEY	KIL	KSD	DTFYNESTFNFKKADDSLND-----983	
Db	1466	ELESSQ	KEARSL	STEL	FLK	NA	YESLEHLETFKREN-----KNLQEBISDLTEQLGSSG 1520	
Qy	984	-----	ESKRR	KLE	BDI	NK	LKKTTL-----QLSPDLNKKYKKLER-LF 1019	
Db	1521	KTIHE	LEKVR	QOL	EAK	ME	LQSALEAEASLEHBECKILRAQLEF---NQIKABIERKLA 1577	
Qy	1020	DKKTVG	KYKM--	OIKL	TLL	KK	QESKLSANPNKHVLQNTSVFPNKKKEAEIAETNTL 1078	
Db	1578	EKDEEM	QAKRN	ILVR	VDS	LQTS	LSLQTSLSLQTSRNEALRV-----KKKMGDLMEMEIQL 1629	
Qy	1079	ENTKIL	KHKY	GLV	KY	YNGE	SSPLKTLSPESIQTDENTYASLENFKVLSKLEGLKMDNL 1138	
Db	1630	SHAN	MAAE	AKQVK-----	SLQSL	LKDTQ	IQLDDAVRANDD-----LKENIAI 1673	
Qy	1139	EKKKLS	YLSGL	HLHL	IAEL	KE	VIKNNKNTGNSPSNNTDV-----NNALESYKKFL-PE 1191	
Db	1674	VERRN	LLQAL	EEL	RAV	VEQ	TERSRLAEQEBIETSERVOILLHSQNTSLINQKKKDDAD 1733	
Qy	1192	GTDVAT	V	W	S	G	S	DTLEQSQPKKPASTHVGAESNITTSQ-----NVDEDDVV 1240
Db	1734	LSQLQTE	VEE	AVQEC	NRABE	EKAKA	ITDAMMAEELKKEQDTSAHLEMKKNKMEQTIKDL 1793	
Qy	1241	IIVP	IFGESE	EYD	DLQGV--	VTGE	AVTPSVIONILSKTENEBEVLYLKLPLAGVYRSKKQ 1299	
Db	1794	-----	QHR	LDE	AEQ	TAL	GKKQKQLOKLARVRELENELE-----AQOKR 1832	
Qy	1300	LENNVMT	FN	VN	VK	DI	LNRSRPNKFNKYNLESDLIPYKDLTSSNYYVKDPYKFLNKBKRD 1359	
Db	1833	NAESV	KMGKRS	ERRIKEL	TYQTE	EDR	KNLNR-----LQDLVKLQKLKVKYKQAEAEAE 1887	
Qy	1360	KF----	LSSYN	YIK	DSID	1373		
Db	1888	QANT	NLSK	FRK	VQ	HELD	1904	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; L15351; AAB59701.1; -;
 DR EMBL; M12995; AAA37081.1; -;
 DR HSP; P08799; IMND.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 782
 FT DOMAIN 783 812
 FT DOMAIN 842 1939
 FT DOMAIN 1439 1443
 FT NP_BIND 178 185
 FT DOMAIN 657 679
 FT DOMAIN 759 773
 FT MOD_RES 129 129
 FT MOD_RES 597 597
 FT MOD_RES 707 707
 FT CONFLICT 1633 1633
 FT CONFLICT 1651 1651
 FT CONFLICT 1686 1687
 FT CONFLICT 1693 1693
 FT CONFLICT 1844 1844
 FT CONFLICT 1879 1879
 FT CONFLICT 1885 1885
 FT CONFLICT 1907 1907
 FT CONFLICT 1928 1928
 FT CONFLICT 1933 1935
 FT CONFLICT 1939 AA; 223626 MW; DBC8297DFE83115A CRC64;
 SQ SEQUENCE

Query Match 4.18; Score 349; DB 1; Length 1939;
 Best Local Similarity 19.4%; Pred. No. 0.00014;
 Matches 297; Conservative 255; Mismatches 577; Indels 400; Gaps 65;

QY 28 LVKLEALEDAVLTCYSLFOKEKVLNNEGTSVATTSTPSKGSVASGGSGSVASGGS 87
 DB 595 LEKNKDLNETVW---GLYQSSSLK-----MATLFST-----YASADAGDSGKGGG 639
 QY 88 VASGGSVASGGSGVSGGGG---NSRRTPNSDSSDSKASVADLKHVRVNYLLTIKELK 143
 DB 640 KKKGSFQTVSALHRENLNKMLTNLRTTHP-----HFVR--CIIPNERK 681
 QY 144 YPQLFD--LTNHLTLCNTHGFKYLDGYEIEINELLYK-----LN-----FYFD 186
 DB 682 APGVMDNPLVMHQLRCNGVLEGIRICRKGFP--NRILYDGRQRYRILNPAAPBQGFID 739
 QY 187 LLRA--KLNDVCANDYCOIPN-----LKRANELDVLK 218
 DB 740 SRKGAELSLSDIDHNQYKFGHTKVFKEKAGLLGLLEEMRDERLSRIITRIQAARGQLM 799
 QY 219 KLVFGYRKPLD-----NIKDNVG-----KMDYTK--KNKKTININEL 255
 DB 800 RIEP--KKNVRRDALLVIQNNIRAFMGVKNWPMKLYFKIRPLKLSAETKEKMANMKEE 857

QY 256 IEESKTKDKNKNATKEEEK--KLYOQYDLSIYNKOLEEAHNLISLVLEKRTDKKNE 313
 DB 858 FCRVSELESKEKSPARKELEKMWVSLQEKNDLQF---QVOAQDNLNDAEERCDQLIKNK 914
 QY 314 -----NIKELDKINEIKNPPANSNTPTLLDRKNKIEEHEKEIKEIAKTIFKNIDSL 368
 DB 915 IQLEAKVKEMTERLEDEE--MNA-----ELTSKKRLED--ECSELKDOI----- 956
 QY 369 FTDPLELEYLREKKNKIDISAKVE-TKESTEPNEYPNGVTYPLSYNDINNALNELNSFG 427
 DB 957 --DDELTL-----AKVEKEHATE-----NKYKNTTEEMAGLD 988
 QY 428 DLINPFDYTKPSKNY-----TDNERKFFINEIKEIKIEKKKIESOKKSYEDRSKSL 481
 DB 989 ELIAKLTREKKAQEAHQALDDDLQAEEDKVNTLTKSVKLE-QQVDDLEGSELEKVKYR 1047
 QY 482 NDIITKEYEKL---LNEIYDSKFNNDIDTINFKMMGKRYSYKVEKLTHHNTFASVENSXK 538
 DB 1048 MDLERAKRKLEGLDNVTOESIMDLENDKLQLE-----EKLKKKEFFDISQONSK- 1095
 QY 539 NLEKLTALKYMEDYSLRNIVVEKELKYKNLISKIENEIETLVENIKKDEEQLFEKKT 598
 DB 1096 -----FEDEQALALQQLKKNQARIEEBEELE--ACRTARAKVEKURSDUT 1142
 QY 599 KDNPKDEKILEVSDIVKQVKVLLMNK--IDELKTKQLILKNVLELKHNIHVNPNSYKOE 656
 DB 1143 RELEISERLEEAGGATSVQIE---MNNKREAEFQKMRDLEEAATLOHEA---TAAAL 1194
 QY 657 NKQEPYLLVLKKEIDKLVFMPKVESLINEEKKNIKTGQSDNSPSTEGEITQATTK 716
 DB 1195 RKKHADSVAELGEQIDNLQ---RVKQKLEKSEKSEFKLE--LDDVTSNMEQIKAKANL- 1247
 QY 717 PQQAAGSALGDSVQAQAQEQKQAOQPPVPVPEAKAQQVPTPPAPVNNKNTVNSKLDYLE 776
 DB 1248 --EKVSRTELDQANRYVKLEESQRSLSNDFTTQRAKLO-----TENGEARQLE 1294
 QY 777 KLYEFLNTSYICHKYLIVSHSTMNE---KILKOYKTKTEESKLS--C 820
 DB 1295 E-----KEALISQLTRGKLSVTQOMEDLKROLEEGKAKNALAHALQASRHC 1342
 QY 821 DPLDLLFNIONNI PMVSMFDSLNLSOLFMEIYEKEMVCLYKLD-NDKIKNLLSEA 879
 DB 1343 DLLREQYEEMBEAKAELQRLVLSKANSEVAQWRTYETAIOITEELEAKKLAORLODA 1402
 QY 880 KKVSTSVKTLSSMQLSLTPQDPKE---VSANDDTSHSTNLNNSLKFENILSGNKK 936
 DB 1403 EEAVEAVNA-KCSSLEKTKHRLQNEIDLWVDVRSNAAAALDKQKRNFDKILAEWKQK 1461
 QY 937 NIYQELIGQKSS-----ENFYEKILKSDSTFYNESFTNFVKSKADDINSL 981
 DB 1462 --YEESQSELESSQKARSLSLTFELKKNAYEESLEHLETKREN-----KNLQBEISDL 1514
 QY 982 ND-----ESKRRKLEEDINKLKKTL-----QLSFDLYNKVK 1012
 DB 1515 TEQLGEGGNVHEKVRKQLEVEKMELOSLAEAEASLEHEEGKILRAQLEF---NQIK 1571
 QY 1013 LKLER-LFDKKTGKTKYM-QIKKLTLLKLEOESLNSLNNPKHVLQNFVSFFNKKKAE 1070
 DB 1572 AEIERKLAEKDEEMQAKRNHLRVVDSLSQTSLEAETRSNEALRV-----KKKMEGD 1623
 QY 1071 IAEENTIENTKILL---KHVGLVLYYNGESSPLKTLSEESIQTEDNYASLENFKVLS 1126
 DB 1624 LNMEIQLSQANRIASEAQHLK-----NAQAH-----LKDTQLQDLDALHAND- 1668
 QY 1127 KLEGLKDNLNLEKKKLSYLSGLHLLIAELKEVLKNKNYTCNSPSENNTDV-----N 1179
 DB 1669 -----LKNIAIVERNTLLQAELEELRAVVEQTERSKRLABQELIETSERVOLLHSQNT 1723
 QY 1180 NALESYKGFPLPGETDVATVWSESGSDTLEQSQPKKPPASTHVGAESNTITTSO----- 1231
 DB 1724 SLINQKKMEADLTQLOTEVEEAQVQECRNAEKAKAITDAAMAEELKKEQDTSALHER 1783
 QY 1232 ---NVDEVDVVIIVPFGSEEDYDLDQGV-VTGEAVTPSVIDNLSKIENEYEVLYLK 1287


```
Qy 550 -----MEDYSURNIVVEKELKYKLNLSKISKEIENEIETIVENIKDEQLEFKEKTIKDNK 603
Db 1093 SOIQSKIEDQALQMOLQKKIKELQARIEEETEE--AERTSRAKAHKRADLSRELEE 1150
Qy 604 PDEKILEVDIVKQVQKVLVANK--IDELKKTOLILKNVELKH----- 645
Db 1151 ISELEEAGGATAAID-----MNRKAEAFQKMRDLBEATLQHEATAAARUKKHADSTA 1206
Qy 646 -----NIHPNSYKQENKQEPYVLIIVKKEID-----KLKVFMPKVESLINEE 688
Db 1207 DVGEQIDNLRVQKLEREKSE-----LKWEIDDLASNMESVSKAKANLEKMCRSLEDQ 1260
Qy 689 KNIKTEGQSDNSESTGEITGQAT-----TPGQOAGSALEDVSQAOAQEQKQAPVPP 745
Db 1261 LSEIKTK---BEEQORTINDISAQARLQTESGEYSRQVEKDALISQSRKQA----- 1312
Qy 746 VPVPEAKAQVPTPPAPVNNKNTENVSCLDYLEKLYEFLNTSYICHKYILVSHSTNMEKILK 805
Db 1313 -----FTQOIEEL----- 1320
Qy 806 QYKITKEESKLSCDPLDLFNIGNIQQNPVMSFD-----SLNLSLSQLFMEI----- 854
Db 1321 --KRHEEIEIRAKKC-PAHALQSARHCDLLREQYEEQEAQKGLQRLSRANSEVAQWR 1377
Qy 855 --YKEMVCNLYKLD-NDKTNLLEBAKKVSTSVKTLSSSSMQPLSLTPQDKPE---VS 908
Db 1378 TKYETDAIQTEELEBEAKKKAQRLQDAEHEAVNS-KCASLEKTKORLQNEVEDLMID 1436
Qy 909 ANDDTSHSTNLNLSLKFENILSLGKNKIYOEL-----IGOKSS-----ENFEY 953
Db 1437 VERSNAACAALDKKQKFDKILSEWKQ--YEETOAELEASQKESRSSTELFKMKNAYE 1494
Qy 954 KILKDSDTFYNESFTNFVSKKADINSND-----ESKRKKLEEDINKL--- 998
Db 1495 ESLDLHLETKREN-----KNLQESLDTQEAEGGKAIHELEKVKQIQEOKSELOPAL 1549
Qy 999 -----KTLQSLFDLYNKYKLLER-LFDKKTGVKQYM-QIKKLTLLKEQLES 1044
Db 1550 EEAASLEHEGKILRVQLEL-NOVKSDIDRKIAEKDEIDQLKRNHLRVVDSMQSTLDA 1608
Qy 1045 KINSLNPNKHVLQNFSPFNKKKEAEI-----AETENTLENTKILLKHYKG 1090
Db 1609 EIRS-----RNEALRLKKMEGDLNEIETQLSHANRQAAEQKLNRTQGVLK--- 1656
Qy 1091 LVKYNGESSPLKLTSESIOTEDNYASLENFKVLSKLEGKLDKNLNEKKKLSYLSGL 1150
Db 1657 -----DTQIHLDDALRSQED-----LKEQVAMVERRANLLQAEI 1690
Qy 1151 HHLIAELKEVIKNKNYTGNSPSENNTDV-----NNALESYKKFLPEGTDVATVVSSEGS 1204
Db 1691 EELRAALEQTERSKVBAEQELLDASERVOLLHTQNTSLINTKKLE--SDISQIQSEM-E 1747
Qy 1205 DTLEQS-----OPKKPASTHVGAESENTITSO-----NVDEVDVDDVLIIPFGES 1249
Db 1748 DTIQEARNAEKAKAITDAAMMAEELKKEQDTSAHLEMKKNLDQTVKDL----- 1798
Qy 1250 EEDYDDLQGV-VTGEAVTPSVIDNLSKIENEYEVLYLKPLAGVYRSLSKOLE--NNVMT 1306
Db 1799 QHRLDEAEQALGKGGKQIQKLEARVREGEVDA-----EQKRSAAEVKGVK 1847
Qy 1307 FNVNVKDIILSRFNKRNFKNVLSLDLIPYKDLTSSNVVVKDPYKFNKERRKDFLSYN 1366
Db 1848 YERRYKEL---TYQSEEDRKNVLR-----LQDLVDKLMQVKYSKQRAEEAE--LSNVN 1897
Qy 1367 YIK 1369
Db 1898 LSK 1900

RESULT 39
SMCL_YEAST
ID SMCL_YEAST
AC P32908;
STANDARD; PRT; 1225 AA.
```

```
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome segregation protein SMC1 (DA-box protein SMC1).
GN SMC1 OR CHL10 OR YFL008W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103320; PubMed=8276886;
RA Struninikov A.V., Larionov V.L., Koshland D.;
RT "SMC1: an essential yeast gene encoding a putative head-rod-tail
RT protein is required for nuclear division and defines a new ubiquitous
RT protein family."
RL J Cell Biol. 123:1635-1648(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-268(1995).
CC -!- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH SMC2 OR OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00602; AAA16595.1; -
CC EMBL; D50617; BAA09230.1; -
CC PIR; A49464; A49464.
CC PIR; S41804; S41804.
CC SGD; S0001886; SMC1.
CC InterPro; IPR003405; SMC_C.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF02463; SMC_N; 1.
CC Pfam; PF02483; SMC_C; 1.
CC
CC Mitosis; ATP-binding; Coiled coil; Nuclear protein.
CC NP_BIND 33 40
CC DOMAIN 173 489
CC COILED COIL (POTENTIAL).
CC COILED COIL (POTENTIAL).
CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC ALA/ASP-RICH (DA-BOX).
CC S->L: IN TS MUTANT SMC1-2.
CC MUTAGEN 173 173
CC MUTAGEN 458 458
CC N->D: IN TS MUTANT SMC1-1.
CC SEQUENCE 1225 AA; 141279 MW; B504017AA0ECCAC8C CRC64;
Query Match 4.1%; Score 347; DB 1; Length 1225;
Best Local Similarity 20.3%; Pred. No. 9.6e-05;
Matches 284; Conservative 248; Mismatches 505; Indels 364; Gaps 69;
Qy 203 IPFNLKIRANEL--DVLKKLVFGYKPLDNLIK-----DNVGMED-----YTK-----KNKK 247
Db 46 ISFVLGVRSNHLRSNLIKDLI--YRGLNDENSDDDYDNEGAASSNPQSAAYKAFYQKGNK 103
Qy 248 TIENINELIESKTTIDKNKNATKEEEKKLYQAOYDLSIYNKO-LEBAHNLIISVLEKRI 306
Db 104 LVELM-----RIISRNGDTSYKIDGKTVSYKDYISIFLENILIKAKNFL-VFQGDV 154
```


Qy	307	DTLKKNENIKELDLKINETIKNPPPPNPPNTWTLDDKWKKEIEHEKETIKETAKTINID	366
Dy	155	EQIAAQPVPV-ELSRMPEEV-----SGS-----IQYKKEYEELKEKTEKLSKATESIK	201
Qy	367	SLFTDPLELEYILREXKNKIDISAKVEYKESTEPNEYPNGVYPLSYNDINNALNSF	426
Dy	202	NRRIHGEUKTYKEGINNVEYKQDKK-----NELQKF	236
Qy	427	GDILNPFDTKPSKNITYD--NERKKFINEIKETIKKKIBSDKKSVEDRS-----	478
Dy	237	QALWQL--YHLEQOQKEELDKLSALNSEISSLKGKINNEMKSLQBSKSFVKESAVISKQ	294
Qy	479	-KSLNDITKEYEKLNEIYDSFNENNIDJTNPEKMMGKRYKSYKVEK-----LTHHNT	529
Dy	295	KSKLDYIFDKELVSDLRILKVP-----QQAAGKRISH-TEKRTIESLQKDLQOKT	345
Qy	530	FASYENSKHLEKTLKALYMED--YSLRNI-----VVEKELKYKKNLSK-----I	574
Dy	346	YV--ERFETQLKVVYTSKAFBEEIKQSARNYDKFKLNENDLTKYNCLHEKYLTEGGSIL	403
Qy	575	ENBIETLVENIKKDEBQL-----FEKTKTDENKPDKEI-LEVSADI-VKVOQVKVL	623
Dy	404	EEXIAVLNNDKREIQEELERFNKRADISKRRITELSITGEKLDQLNDLRVSLNEKNAL	463
Qy	624	LMKKIDELKQTQILKLNKELNINIIVPNSYK-----QENKQPPYILVLKKEID	672
Dy	464	HTERLHELAKLO--SDIESANNQEVDLNFKRLRETLVKIDOLSANQRETMKERKLENIA	520
Qy	673	KLAVEMPKVESLINEEKNIKTEGOSDNSEPSTEGEITQOATTPGOQAGSAL--EGDSV	730
Dy	521	MLARFPFGVGLVH-----DCHPKKE-----KYGLAVSTILKGNFDS	559
Qy	731	QAO-----AQE-----OKAOQPPVPVPPEAKAQAQVPTPPAP--VNNKTENVSKLDYLEKLY	779
Dy	560	IVENLTVAOECIAFLKQKAGTASFIPLDTIETELPTLSLPSQDYILSINAIDY-EPEY	618
Qy	780	EPLNTSYCHUYILVSHSTWNEKILKQYKITHKEESKLSLSCDPLDLFNTONNTPVMSY	839
Dy	619	E-KAMOYVCGDSIIC--NFTLN--IAKDLKWKKGIRGLVITIE-----GALIHK	661
Qy	840	FDSLNNLSQLFMPEIYEKKEWCVNLVKLDN-----DKIKNLE-----EAKKYSTSVKTL	890
Dy	662	AGLWTGGISCDANNRWKDEEYQSLMSLKDQLIQIDELNSGORSNTPAREVENSVLIN	721
Qy	891	SSSQPLSLTPQDKPEVSANDTSHSTNNLSKLFIENILSGKNNKIYQELIGQKSS--	948
Dy	722	SDIANLRTOVTOOK-----RSLDENRLRLEIKYHNDLIEKIQPKITEL	763
Qy	949	-----ENFYEKILADSDTFYNESFTNFVKSADDDINSLNDESKR--KLEEDINKLK	999
Dy	764	KKLDLLENTKMLVKEKALQNNIFKETPSKIGTTFKEYENHSGELMROOSKLELOLQK	823
Qy	1000	TLQLSFDLVNKKYLKLERFDKKTGVGYKMQIKKLTLLKQLESKLSLNNKPNHVLQNF	1059
Dy	824	QI--LTVENKLOFETDRL--STTORRYEKAQKL-----ENAOVEMKSLBEEQVAI--	870
Qy	1060	SVFNNKKKEABTAETENTLIENTKILKHVGLVYKNGESSPLKTLSPESIQTEDNYASL	1119
Dy	871	-----EMKIGSIESKLEEHK-----NHLDELQKFVTKQSELN--SSSEDILEDMN---	913
Qy	1120	ENFKVLKLBGLKLDNLNLEKKKLSVLSGLLHLIAELKEVINKKNYTG-NSPSENNTDV	1178
Dy	914	SNLOVLKRRDGLKE--DIEKFDLBERVA-----LKNCIKSNINIPISSETTI	959
Qy	1179	NNALESYKXKFLP-EGTDVATVYSESGSDTLEOSQPKPASTHVGASNTITTSQNVDDV	1237
Dy	960	DD-----LPISSTDNEAIIISINIDINYKGLPKK-----YKENNTDSARLELQKI	1005
Qy	1238	DDVIIIPFGESE-----EDYDDLJ--QVVTGE-----AVTPSVIDNLSKINIEYEV	1283
Dy	1006	HEV--BEILNEQPNARALRYDEAGREVFVNNNETEOLKAEEKKILNOFKLKKRKE	1062

[illegible]

Db 5 KLE-----VRGFTYKKPQVI-----DFTF-LKFFVIQKGTG---AGKTSIID 43
Qy .91 GGSVASGSGSRRTPNSDSSDAKSYADLKHVRNLYLTIKELKYPQ--- 146
Db 44 AITVALYGVPRYGA--SVATKYVLSRGEKELKVALDFSLGRNRYRERYREFPEDSOV 101
Qy 147 -LFDLTNHLMTLCDNIHGFYKLYIDG--YEEINELLYKLNFYFOLLRAKLDNDVCANDYCOI 203
Db 102 RVEEGRRLNIKANEVEKWLKFKISGLDYKTKVILLPQGEFDFLKE-----SSERKKI 156
Qy 204 PFNLKIRANELDVLKLVGVRKPLDNKDNVGK-----MEDYIKKN-----KTIENINEL 255
Db 157 LINE-LGLELEKVRQLASFTFKNLEKREALKREYELLKDYTPTRKVELEKTLKNLEE 215
Qy 256 IEESKTKTDKNKATKEEKKKLYQAOVDLSYINQKLEEAHNLISVLEKTRIDTLKN--- 312
Db 216 LKELTEEKLRLQELKKAEDK--SLRELQSVYTKLELN-----LEKEVEKLEKLEF 269
Qy 313 -----ENTKELLDKINEIKNPPPPANGTNPNTLDKNKKIEEHEKEIKEIAKTI 361
Db 270 SRKVAPYVPIAKRIEEDKKLTLEK-----VRKNKL-----TKELAVLKDEL 311
Qy 362 KFNIDSLFTDPLEYLYLREKNKNIDISAKVETKESTEPNPGVTPYPLSYNDINNALN 421
Db 312 SFAQELNLRIDAEKEKEKEKEKEKELEHRLK-----KLOEIKELK 353
Qy 422 ELNSFGDLINPFDYTKPSKNIYTD-NER-----KKFNEIKEKI-KTEKKKIESDKKSYE 475
Db 354 ELSQSSSLAKEKEEYEQAKQEFEDLSERVEKGLKVAETEELKLEKLEFSEEEYSLK 413
Qy 476 DRKSLNDITKEYEKLKLEIYDSEKFNNDILTNPEKMGKRYSKYKVEKLTHHTTFASYEN 535
Db 414 MKERLLVELQRLKLEL-----KEKEGOLENLTQ-----KYKE 445
Qy 536 SKHNLKLTALKWEDYSLRNIVVEKELKYKLNLSKIENEITF-----LVEN 584
Db 446 KKKVHEKVLNLEKELE-----RELKERLHYHAHVASYLSPGDTCPVCGGIYRGKALEN 500
Qy 585 IKKDEEQLFEKKITKDENKPKDEKILEVSDIVKVOVKV-----LLMKNIDELKKTOL 636
Db 501 V--DAEGISELKHAKELKEER--EIDTTLKLYAQKINSLEKEMEKLRNEVEELK--- 553
Qy 637 ILKNVELKHNTHVPSYKQENKQEPYLLVLKLEIDKLVKMPKVESLINESKKNIKTEG 696
Db 554 -----EIPENLKER-----IKKLEELRIEKEKLEHKLKNYK----- 585
Qy 697 QSDNSEPSTEGITQATTKPGQAGSALGDSVQAOQEOQKQAPVPVPVPEAKAQVP 756
Db 586 -----ALE-----DRQKQK-----BAQAKL- 601
Qy 757 TPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHKYLVSHTMNEKILKQYKITEESK 816
Db 602 -----HKAQTELEL-----LKEKREKSLRVEKPE- 627
Qy 817 LSSCDPLDLQNIQNPVMYSMFDSNLNSLSQLFMEIYERKMCNLYKLKD-----NDK 871
Db 628 -----LYRVER-----LEDYEE-----SLKEEINYNSK 651
Qy 872 IKNLLEBAKVSTSVKTLSSSSMQPLSTPODKPEVGSANDTSTNLSNLSKLFE--NI 929
Db 652 LQETEEKKKLRKHFEESSRSKSL-----EGELSALNESINSLEESERKEKLELANI 704
Qy 930 LSLGKN-KNIIQELIGOKSSSENFYEKILKQSDTFYNESFTNFVKSADDSINLNDSEK 988
Db 705 YEVAKSPEVVELYLGQKAE--LERKIKE-----FEESFQSL-----KLKKS 745
Qy 989 KLEEDINKLKTQLSFDLYNKKYKLEKLERLFDKKTKYKQWKQIKKLTLLKEQLESKINS 1048
Db 746 EIEEKLKEYEIGRELS-DIKEYESVKTOLKEKHKKLGEVK---RELEHLGERLKR--- 798
Qy 1049 LNNPKHVLQNFVFPFNKKKAEIAETENTLTKILKHKYKG-LVKYNGSGSSPLKTL-- 1105
Db 799 -----BELQKEISELEKKLEVYRVISNDFRGDRFQKYVSEIMLQKVVD 842

Qy 1106 -SEESIQTEDNYASLENFKVLKLEGLKONLNL-----KKKLSYLSGG 1149
Db 843 ASEFYFTGNYF-----FELERATKGRDKDIVVESSTSQRRPVSSLSGG 888

RESULT 41

MYH7_MESAU STANDARD; PRT; 1934 AA.
ID MYH7_MESAU STANDARD; PRT; 1934 AA.
AC P13540; Q60540;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
GN MYH7.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIB; TISSUE=Liver;
RX MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
heavy chain gene from Syrian hamster";
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 962-1935 FROM N.A.
RX MEDLINE=88247788; PubMed=3380703;
RA Jandreski M.A., Sole M.J., Liew C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
heavy chain";
RL Nucleic Acids Res. 16:4737-4737(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEATITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; L12104; AAA62313.1; -;
CC EMBL; X07273; CAA30256.1; -;
CC PIR; A28298; A28298.
CC HSP; P08799; 1MND.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR004009; Myosin_N.
CC InterPro; IPR002928; Myosin_tail.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00063; myosin_head; 1.
CC Pfam; PF00612; IQ; 2.
CC Pfam; PF01576; Myosin_tail; 1.
CC Pfam; PF02736; Myosin_N; 1.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS0096; IO; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 779 MYOSIN HEAD-LIKE.
FT 780 809 IQ.
FT DOMAIN 839 1934 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 554 676 ACTIN-BINDING.
FT 756 770 ACTIN-BINDING.
FT MOD_RES 128 128 METHYLATION (SH-1) (POTENTIAL).
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 966 966 D -> E (IN REF. 2).
FT CONFLICT 978 978 T -> TE (IN REF. 2).
FT CONFLICT 986 986 E -> Q (IN REF. 2).
FT CONFLICT 1008 1014 DLQAEED -> ALEARKT (IN REF. 2).
FT CONFLICT 1057 1057 D -> Y (IN REF. 2).
FT CONFLICT 1060 1060 L -> V (IN REF. 2).
FT CONFLICT 1095 1095 D -> N (IN REF. 2).
FT CONFLICT 1217 1217 E -> D (IN REF. 2).
FT CONFLICT 1271 1271 D -> N (IN REF. 2).
FT CONFLICT 1327 1327 T -> A (IN REF. 2).
FT CONFLICT 1358 1358 C -> R (IN REF. 2).
FT CONFLICT 1504 1504 L -> V (IN REF. 2).
FT CONFLICT 1537 1537 M -> L (IN REF. 2).
FT CONFLICT 1556 1556 N -> K (IN REF. 2).
SQ SEQUENCE 1934 AA; 222928 MW; FDBAC59310B0B57D CRC64;

Query Match 4.1%; Score 344.5; DB 1; Length 1934;
Best Local Similarity 19.4%; Pred. No. 0.0019;
Matches 290; Conservative 259; Mismatches 593; Indels 353; Gaps 62;

QY 47 QKEKVLNKGSTGTVTTTPGSKGVSAS-GSGGVSASGSGVSASGSGVSASGGS 105
DB 594 QNKQDPLNVTVGLYQKSKLLSNLFANYAGADAPVDKGGKAKKSGSFQTVLHREN 653
QY 106 GNSRRTNPDSNDSADKSYADLKHRVNYLLTIKELYPQLFD--LTHRMILTCDNIHG 163
DB 654 LNKMLNLRSTH-----PHEVR--CIIPNETKSPGVNDPLVMHQLRCNGVLGS 700
QY 164 FYKLDGEEINEELLYK-----LN-----FYFDLLRA--KLNDVCANDYQIIPN 206
DB 701 IRICRKGFP--NRILYGDPRQRYILNPAAPGQFIDSRKGAELSLSLDIDHNQYKFG 758
QY 207 -----LKIRANELVDLKKLVFGYRKPLD-----N 230
DB 759 HTKVFFKAGLLGLEEMRDLRSRIITRIQASRGLLSRMEF--KKLLERRDLSLVIQWN 816
QY 231 IKDNVG-----KMEDYIK--KNKKTIENTINELIEESKKTIDKNKNATKEEEKK- 276
DB 817 IRAFMGVKNWPMKLYFKIKPLKSAETEKEMATKEEFGRYKDALEKSEARRKELEEKM 876
QY 277 -KLYQAYDLSTYNNKQLEAHNLISVLEKRDITLKNKNE-----NIKELLDKINEKNPPP 330
DB 877 VSLLOEKNDLQL---QVQAEQDNLADEERCQDLKKNKIQLEAKYKEMTERLEDEE--- 929
QY 331 ANSGNTPTNLLDNKKKIEBEKEIKEIAKTIFKNIDSLFTDPLELYLYLREKNKNIDISA 390
DB 930 -----EMNAELTAKKRLEDECESELRDID-----DLELTL-----A 961
QY 391 KVE-TKESTEPNEYPNGVTYPLSYNDINNALNELNSFGDLINPFDYTKPESKNIVTDNER 449
DB 962 KVEKDKHATE-----NKVKNLTEEMAGLDEIIAKL--TKE----- 994
QY 450 KKFINEIKKIKIEKKIESDKSYEDRSKSLNDITKEYEKLILLNEYDSKFNNNIDLTNFW 509
DB 995 KKALQEAHQO-ALDDILOAEEDK-----VNTLTFSKVKLEQQVDD--LEGSLQEKK 1042
QY 510 EKMGMKRYSKYE---KLTHHTTFASYENSKHNL-EKLTGK-----ALK-YMEDYSLRNI 558
DB 1043 VRMDLERAKRKLEGLDKLT-QESIMDLENDKQOLDEKLLKKDFELNALMARIEDEQALGS 1101

QY 559 VVEKELKYKNLISKIENIEIETLVENIKKDEQLPEKKITKDKENKPEKILEVSDIVKQV 618
DB 1102 QLOKKLKELOARIBELEBELE--AERTAKVKEKLRSDLSRELEISERLEEAGGATSVQ 1159
QY 619 VOKVLLMNK--IDELKTKTQLILKYNVELKHNIHVPSYKQENKQEPYLYLVLKKEIDKLKV 676
DB 1160 IE---MNMKREAFQKMRDRLEEAATLQHEA---TAAALRKKHADSVAELGEQIDNLQ- 1210
QY 677 FMPKVESLINEEKKNIKTGEGSDNSEPSTGEITQAT-----TKPGQAGSALGDSV 730
DB 1211 ---RVQKLEKEKSEFKLE--LDDVTSNMEQIIKAKANLEKMCRTLEDOMNHRSAEET 1265
QY 731 QAQAQEQKQAQPPVPVPPVPEAKAQAQPTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHK 790
DB 1266 QRSVNDLTSORAKLQTENGELSRLDEKALISQLTR--GKLTYYQOLEDL----- 1314
QY 791 YILVSHSTWNEKILKQYKITKEESKLSCLDPLDLNFIQNNIPVMSFDSLNNLSOL 850
DB 1315 -----KQLEEEVKAKNTLAHALQSAHRDCDLLREQYEETEAKABLCQVLSKANSEVAQ 1369
QY 851 FMEIYEKEMVCNLYKLD--NDKIKNLLBEAKKVSTSVKTLSSSSMOPLSLTPQDKPE--- 906
DB 1370 WRTYETDAIQTETELEEAKKLAORLQDAEAEAVNA--KCSSLEKTKHRLONEIEDLM 1428
QY 907 VSANDDTSHSTNLNNSKLFIENILSGKNKNYQELIGQKSS-----ENF 951
DB 1429 VDVERSNAALAAKQKRNFDKILAEWKQ--YEESQSELESSQKEARSLSTELFKLNA 1486
QY 952 YEKILKDSDTFYNESTFNVKSKADDDINSND-----ESKRRKLEEDINKLAK 999
DB 1487 YEESEHLETFKREN-----KNLQEEISDLTEQLQSTGKSIHEIKIRKQAEKMEELS 1541
QY 1000 TL-----OLSFDLNRYKYLKLER-LFDKKTVTGKYM-QIKKLTLLKE 1040
DB 1542 ALAEAASLEHEGNNILRAQLEF---NQIKAEIERKLAEKDEMEQAKRNHLRVVDSLQ 1598
QY 1041 QLESKLSLNNPKHYLQNFVFNNKKKEAEIAETENTLTKILLKHYKGLVYKYNNGESS 1100
DB 1599 SUDAETRSRNEALRV-----KKMEGDLNEMEIQLSHANRMAAEQAQKQV-----S 1644
QY 1101 PLKLTSEESIoTEDNVASLENFKVLSKLGKLDNLNLEKKKLSYLSGSLHLHIAELKEV 1160
DB 1645 LOSLKDQTOIQDDAVRANDD-----LKENIAIVERNNLLOAELEELRAVVEQT 1694
QY 1161 IKNNYVTGNSPENNTDV-----NNALESYKFL-PEGTDVATVYVSESGDTLEQSQPK 1213
DB 1695 ERSRLAEQELIETSERVQLLHSHQNTSLNQKKMDADLSQLQTEVEEAQVQECRANAEKA 1754
QY 1214 KPASTHVGAESNTITTSQ-----NVDDEVDDVLIIVIFGESEEDYDDLQGV-VT 1261
DB 1755 KKAITDAAMMAEELKEQDTSAHLERMKNMEQTKDL-----QHRLDEAEQIALK 1805
QY 1262 GEAVTPSVIDNLTISKENEVEVLYLKPLAGVYRSKLGKOLENNVMTFNVVVKDILNSRFNK 1321
DB 1806 GKKQLQKLEAVRELENELE-----AEQKRAESVKGMKRSERIKELTYOT 1853
QY 1322 RENFNKVLSDLPYKDLTSSNVVVKDPYKFLNKEKRRKF---LSSYNYIKDSID 1373
DB 1854 EEDRNLLR----LQDLVDKLQKLVKAYKQAEAEAEQANTNLSKFRVQVHOLD 1903

RESULT 42
YCF2_MARPO
ID YCF2_MARPO STANDARD; PRT; 2136 AA.
AC P09975;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 259 kDa protein ycf2 (ORF 2136).
GN YCF2.
OS Marchantia polymorpha (Liverwort).
OG Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=3197;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
RA Ozeki H.;
RT Liverwort Marchantia polymorpha chloroplast DNA.;
RL Nature 322:572-574(1986).
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04465; CAA28078.1; .
DR PIR; A05037; A05037.
DR PIR; S01591; S01591.
DR InterPro; IPR003959; AAA_ATPase_centre.
DR Pfam; PF00004; AAA; 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match 4.1%; Score 344; DB 1; Length 2136;
Best Local Similarity 19.8%; Pred. No. 0.00022;
Matches 338; Conservative 243; Mismatches 541; Indels 584; Gaps 80;

Qy 160 NIHGFKYLDIGYEINELLYKLN---FYF-----DILRAKLD-----VC 196
Db 204 NIDELKHFEEY---INQYIPNNWMEYFYFIFINQLKIDIKSNKNSIGFEVFLAFC 261
Qy 197 ANDYQIPNLIKIRANELDVLKLVGYRKPLDNDKDNVGMEDYI---KK---NKKTI 249
Db 262 EKLLEFEVFLSKPNNNLOM--KL-----NCLNFSFELDFICILNKKLPWVNKKIF 310
Qy 250 ENINELIESKTTIDKKNKATKEBEKKLYOQYDLSIYNNKOLBEAHLNISVLEKRIDTL 309
Db 311 KNLQNFESDKLIE-----SFFLLKIKGNLYFKNYIEFTVWQSYKKDCLDFN 358
Qy 310 KKNE-NIKELLDKINEIKNPPAN-----SGNTPNTLLDNKKIEHEKEIEIARTI 361
Db 359 KFNELNNEIYKIEELFSDIYKFSKYLVEGKKRTIHKOSFNNNNIYKKLNSI----- 414
Qy 362 KFNIDSLF-----TDPLEYLYLREK-----NKN 385
Db 415 -FNFTIYPDPSNLLFDWLKKNYINNKPFLKSLIYSSISNQIFLFFKQKNSKFNKN 473
Qy 396 I-----DISAKVETKS-TEPNYVNGVTYPLSYNDINNALNELNSGDLINPDYTKE 438
Db 474 LVKNSKDVITNVSKEKIEINNFSKIY-----AFFELISNEIDNKFVINKI 524
Qy 439 PSKNIYTDNERKKFNEIKER-----IKIEKKIESDKKSDKSDRSKLN---DITKE--Y 488
Db 525 SLKNNKKQKRFYLNKISSDNFRFINLWKIKYSSQOFVSNNSFLINPAFELQQNY 584
Qy 489 EKLNEIYDSEFNNDITNFKMMGRYSYKVEKLTHTNFASVEN--SKHNLKLTKA 546
Db 585 LKKKNILFFKLNE-----VPSNFEYQY-YKCKNLNIFLAFASLEKILKKRN-KKFTIS 637
Qy 547 LKYMEDYSLRNVVKEKLYKYNLISKIENEIETLVENIKKDEQOLFETKTKDENKPE 606
Db 638 IKLEF-----KFKYKNLN--ENGEYKIESQILQNEKELNKKRKNQFQNPNI 682
Qy 607 KILEVSDIVKQV-----QKVLNKKIDELKK--TQLLKNVEL-----KHNHVP 650
Db 683 KILSFYSSKKNIYLNQYFFKNKLNKLTITWKIKSNKLVISSEYNNKIIWNKKMKFF 742

Qy 651 NSYKOE-----NKQEPYIYLIVKKEIDKLVKVPKVESLIN-----EEKKNITEGOS 698
Db 743 SFSKNSVLDTFFFNKSKFNILITVIFDKLKKIQLNFOEIQILNCFSLFNSKNIKTKIF 802
Qy 699 DNSEPTSEIGTQATTKPGQOAGSALBGDSVQAQAQEQQAQPPVPVPEAKAQVPTP 758
Db 803 KNSYFINEN-----LTT 814
Qy 759 PAPVNNKTENVSKLDYLEKLYEFLNTSYICH---KYILVSHSTMNEKILKQYKITEEES 815
Db 815 TFSNDKREFNIP---FLELFISEINNDFLMRFKKYL-----YRIYKDEI 858
Qy 816 KLSSCDPLDLFNIGNNIPVMSFDSLNN---SLSQLFMEIYEREMVCN---LYK--- 865
Db 859 LFNPIENQOLLNPEKTKIL-TFIDFLQDPELVNRRFIFHLEKTKIKNNLLLYRLLK 917
Qy 866 --LKDN-----DKTNLLEBAKVKSTVKTLSSSSMOPLSLTPQDKPEVSANDTSHST 917
Db 918 IFLDKRNFLLINEIKSFIE--KKNLFIKLSQLSNVL----- 953
Qy 918 NLNLSKLIFENITLSL---GKNKIYQELIGOKSSENYEK----- 954
Db 954 -VKNSYKFFDNIFNPHFLKQKEKNI--EII--LNNQNTFEKSLKTKYLKLNLNNSYSK 1008
Qy 955 -----ILKDS-----TFYNESPTNFVSKADDIN---SLNDESKRKK 989
Db 1009 PSYKIFIFOLLNILNKNYKTFOWISELIFYSKNLNYKIQNKIENNNYCNKKNISYK 1068
Qy 990 LEE-----DINKLKT-----LQSLFDLYNKYL-----KLERLEDFKKTGVK- 1027
Db 1069 IKTWNFEKNNLFQTNNSWFTLEWMEYNTVLLQIQETFFQIDVLVEYKPKKIIIEKN 1128
Qy 1028 -----YKMOIKKTLTLKEQLESKL---NSLNNPKHVLQNFVFFNKKKEAIEAETNT 1077
Db 1129 LKFFLKSKKISLKTLSFHNFKLWNLRFFNEINYKKNLLNF-----LWSDFN 1177
Qy 1078 LENTKILLKHYKGLV---KYVNGESSPLKLTSESIQTEONYASLENPKVL----- 1125
Db 1178 INNCCNLYWVLSLVIFIFLYQ-----KIFS---IIIGSCDFHLWKKFIIQYLTDRSR 1229
Qy 1126 ---SKLEGKLDKNLNEKKKLSYLSGLHLHLIAELKEVINKNKYTNGSPSENNTDVNNA 1181
Db 1230 SLYFTKLTTRNKLTALNKTENLLSYFQNLTHYITNIKPYLLTK----- 1272
Qy 1182 LESYKKELPECTDVATVVSSESGDTLEOSQPKPASTHVGAESNTITTSQNVDDDEVDD-- 1239
Db 1273 -KNLKKWL-----INNKTLDJUSRRRKKLLVOSLTHNKI---QNTGFELSNK 1316
Qy 1240 VILVPIFGESBEDYDDLQGVVTGEAVTPSVIDNILSKIENEYEVLYLKLPGVYSLKKQ 1299
Db 1317 QFTSYFG-----YOITNOOGLLYFOYLAQFF---QKN 1346
Qy 1300 LENNVN-----TFNVNWKDILNSRFNKNR-----FNKV---LESDLIPYKDL----- 1339
Db 1347 LINSLDLANKWIVFSEWHKIFSOKLRQTNIELGFQNIIPVLPQGLSYSKGILLIGPI 1406
Qy 1340 -TSSNVYVKD-----PYKF-LNK-----EKDRKFLSSY----- 1365
Db 1407 ETGRSYLILKNLAESYVPLFKISINKNLYNKPDPVITESMMNIIIESLRNLNLTDFAKKM 1466
Qy 1366 -----NYIKSDITDINFANDVLYGYKILSEKYSKSDLSIRK---YINDK 1407
Db 1467 SPCTIWIQNIHQLVNRLTQNVESDPTF---LLG---ILLKYFQDFSKTKNNIIVIGS 1520
Qy 1408 QGENEYLPFLNNIETLYKTVNDKIDLVIIHLEAKLVNTYEKSNVEVK----- 1456
Db 1521 THLPKKVDPALISPNRLDKIIN--VRLENI SORKKQFPLLKLLKKNFOLKNFLFNERS 1578
Qy 1457 -----IKELNYLTIQDKLADFKKNNFVGIADLSTQYNH-----NLLTKFLSTG 1502
Db 1579 RTMGYNLRDL SAL-TNEVLLISITKNRSFIDTDLTKLAFHFRQIFGLTYTNNKLNDFRIFK 1637

QY 1503 MYFENLAKTVLSN-LLDGNLQGLMNI 1527
ID : : : : :
Db 1638 IVIYKVGKTIQNILIKSSMNLNI 1663

RESULT 43

MYH6_HUMAN STANDARD; PRT: 1939 AA.
AC P13533; Q13943; Q14906; Q14907;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
GN MYH6 OR MYHCA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92133665; PubMed=1776652;
RA Matsuoka R., Beisel K.W., Furutani M., Arai S., Takao A.;
RT "Complete sequence of human cardiac alpha-myosin heavy chain gene and
RT amino acid comparison to other myosins based on structural and
RT functional differences.";
RL Am. J. Med. Genet. 41:537-547(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94140346; PubMed=8307559;
RA Epp T.A., Dixon I.M., Wang H.Y., Sole M.J., Liew C.-C.;
RT "Structural organization of the human cardiac alpha-myosin heavy
RT chain gene (MYH6).";
RL Genomics 18:505-509(1993).
RN [3]
RP SEQUENCE OF 1-177 AND 1551-1939 FROM N.A.
RX MEDLINE=89264452; PubMed=2726733;
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RT "Characterization of human cardiac myosin heavy chain genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
RN [4]
RP ERRATUM.
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
RL Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
RN [5]
RP SEQUENCE OF 1407-1939 FROM N.A.
RX MEDLINE=88299163; PubMed=2969919;
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
RT "Molecular cloning and characterization of human cardiac alpha- and
RT beta-form myosin heavy chain complementary DNA clones. Regulation of
RT expression during development and pressure overload in human
RT atrium.";
RL J. Clin. Invest. 82:524-531(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D00943; BAA00791.1; -;
DR EMBL; Z20656; CAA79675.1; -;
DR EMBL; M25140; AAA60386.1; -;
DR EMBL; M25162; AAA60386.1; JOINED.
DR EMBL; M25142; AAA60387.1; -;
DR EMBL; M25141; AAA60387.1; JOINED.
DR EMBL; M25164; AAA36344.1; -;
DR PIR; A46762; A46762.
DR PIR; A28908; A28908.
DR HSSP; P08799; LMND.
DR Genew; HGNC:7576; MYH6.
DR MIM; 160710; -;
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 842 1939 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP.
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 88 88 Q -> E (IN REF. 1).
FT CONFLICT 574 574 Q -> P (IN REF. 1).
FT CONFLICT 608 608 A -> G (IN REF. 1).
FT CONFLICT 744 744 T -> A (IN REF. 1).
FT CONFLICT 790 790 M -> I (IN REF. 1).
FT CONFLICT 1014 1014 V -> A (IN REF. 1).
FT CONFLICT 1021 1021 S -> T (IN REF. 1).
FT CONFLICT 1101 1101 A -> V (IN REF. 1).
FT CONFLICT 1290 1290 A -> S (IN REF. 1).
FT CONFLICT 1373 1373 W -> C (IN REF. 1).
FT CONFLICT 1533 1533 K -> N (IN REF. 5).
FT CONFLICT 1540 1540 L -> M (IN REF. 5).
FT CONFLICT 1577 1578 KL -> NV (IN REF. 5).
FT CONFLICT 1705 1706 EQ -> DR (IN REF. 1).
FT CONFLICT 1733 1733 E -> D (IN REF. 1).
FT CONFLICT 1734 1734 A -> S (IN REF. 2).
FT CONFLICT 1737 1737 T -> S (IN REF. 1).
FT CONFLICT 1763 1763 D -> H (IN REF. 1).
FT CONFLICT 1788 1788 M -> I (IN REF. 3).
FT CONFLICT 1871 1871 D -> N (IN REF. 5).
FT CONFLICT 1882 1882 R -> G (IN REF. 5).
FT CONFLICT 1890 1890 Q -> R (IN REF. 5).
FT CONFLICT 1933 1933 MISSING (IN REF. 5).
SQ SEQUENCE 1939 AA; 223689 MW; ECB87E7CEB768B6F CRC64;

Query Match 4.1%; Score 343; DB 1; Length 1939;
Best Local Similarity 19.7%; Pred. No. 0.00021;
Matches 304; Conservative 267; Mismatches 542; Indels 432; Gaps 70;

QY 28 LVKLEALEDAVLTVSLFQKMKVNLNEGTSGTAVTTSTPGKSGSVASGSGSVASGGS 87

CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; M76598; AAA37159.1; -
DR EMBL; M76599; AAA37160.1; -
DR EMBL; M76600; AAA37161.1; -
DR EMBL; M76601; AAA37162.1; -
DR EMBL; M62404; AAA37424.1; -
DR HSSP; P08799; IMND.
DR SWISS-2DPAGE; Q02566; MOUSE.
DR MGD; MGI:97255; Myhca.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODOM; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding; Polymorphism.
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP.
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT DOMAIN 790 807 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 816 833 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 129 129 METHYLATION (TR1) (POTENTIAL).
FT MOD_RES 697 697 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 707 707 ALKYLATION (SH-2) (POTENTIAL).
FT VARIANT 194 194 Y -> D.
FT VARIANT 545 545 S -> A.
FT VARIANT 838 838 I -> S.
SQ SEQUENCE 1938 AA; 223564 MW; EAD789ADA68818FB CRC64;

Query Match 4.1%; Score 341.5; DB 1; Length 1938;
Best Local Similarity 20.0%; Pred. No. 0.00024;
Matches 251; Conservative 234; Mismatches 467; Indels 303; Gaps 56;

QY 237 KMEVDYK--KNKKTENINELTESKKTIDKNKATKEBEK--KLYOAYDLSTYNKOL 292
DB 837 KIRPLKSAETKEANKEEGRVDALEKSEARKEEKEKWSLLQEKNDLQ---QV 893
QY 293 EBAHNLISVLEKRIDTLTKNE-----NIKELDKINEIKNPPANSNGNPTNLLDKNKKI 347
DB 894 QAEQNDLDAEERCQOLIKNKIQLEAKYKEMTERLEDEEE--MNA-----ELTAKKRKL 945
QY 348 EBEKEIKIATIKFNDSLTDPLDLELYLREKNKNIDISAKVE-TKESTEPNYPNG 406
DB 946 ED---ECSELKKDI-----DQLELT-----AKVEKEKHATE----- 974
QY 407 VTYPLSYNDINNALNELNSFGDLINPFDYTKPSKNYTDNERKKFNIKEIKIEKKK 466
DB 975 -----NKVKNLTEEMAGLDEIAKL--TKE-----KRALQEAHQO--ALDDLQ 1013

QY 467 IESDKSKSYEDRSKSLNDITKEYEKLNLNEYDSKFNNNIDLTNPEKMMGRKRYKVE---K 523
DB 1014 AEEK---VNTLTKSVKLEQQVDD--LEGSLQEKKKVRMDLERAKRKLEGLDK 1062
QY 524 LTHHTTFASYENSKHNLEKLTALKY-----MEDYSLRNIVVEKELKYKYLNLISKIE 575
DB 1063 LT-QESIMDLNKLQLEKLEKKKEFDISQQNSKIEDEQALALQLOKLEKNQARIEELE 1121
QY 576 NEIETLVENIKKDEQLPEKTKTDENKPKDKILEVSDIVKQVQKVLMMNK--IDELKK 633
DB 1122 EELE--AERTARAKVEKRLSDLSRELEISERLEAGGATSVQIE---MNRKREAFQK 1175
QY 634 TQLIILKNVELKINIHPVNSYKQENKQEPYILVLKLEIDKLVKFMKPVESLINEEKNKIK 693
DB 1176 MRDLEEAATLQHEA---TAAALRKKHADSVAGELGQIDNLQ---RVQKLEKEKSEFK 1227
QY 694 TEGQSDNSEPSTEGETGQATTGPGQAGSALGEGSVQAQAEQKQAPVPVPVPEAKA 753
DB 1228 LE--LDDVTSNMEQITKAKANLEKYSR-----TLEQANEYR-----VKLEAQK 1270
QY 754 QV---PTPPAPVNNKTENVSKLDYLEKLYEFLNTSYICHKYLIVSHSTWNE---KILQ 806
DB 1271 SUNDFTTORAKL--QTENGELARQLEE-----KEALISQLTGKLSYTOQMED 1316
QY 807 YKITEEESKLS--CDPLDLLFNQNNIPVMYSPDLSLNNLSLSOLFMEI 854
DB 1317 LKQLEEGKAKNALAHALQSSRHDCDLLREQVEEEMEAQELQRLVLSKANSEVAQWRTK 1376
QY 855 YEKEMVCNLYKLD-NDKIKNLLEBAKYSTSVKTLSSSSMQPLSLTPQDKPE---VSN 910
DB 1377 YETDAIQRTTELEEAQKLAQLODAEAEAVNA-KCSSLEKTKHRLQNEIEDLMVDVE 1435
QY 911 DDTSHSTNLNLSIKLFENILSLGKNKIYQELIGKSS-----ENFEYKI 955
DB 1436 RNAAAAALDKQKORNFDKILAWEKOK--YESQSELESSQKEARSLSLTFKLKNAYES 1493
QY 956 LKSDTFYNESPTNFVSKKADDINSIND-----ESKRKKLEEDINKKTL-- 1001
DB 1494 LEHLETFKREN---KNLQEBISDLTEQLEGEGKNVHLEKIRKQLEVEKLEQALKEE 1548
QY 1002 -----QLSFDLYNKYKLEKLER-LFDKKTGVGKYM-QIKKLTLLKQLES 1044
DB 1549 AEASLEHEGKILRAQLEF---NQKAEIERKLAEDMEQAKRNHLRWDSLOTSLDA 1605
QY 1045 KLNLSNPNKHVLQNFVSFFPNKKKEAEIAETENTLENTKILL---KHYKGLVKYYNGESS 1100
DB 1606 ETRSRNEALRV-----KKMEGDLNEMEIQLSOANRIASEAQKHLK-----NSQAH 1651
QY 1101 PLKTLSEESIQTEDNYASLENFVKLSKLEKLDNLSNLEKKLSYLSGLHHLIAELKEV 1160
DB 1652 ---LKDQLOLDVAVHANDD-----LKENIAIVERRRNLLQAELELRAYVEQT 1697
QY 1161 IKKNYVTGNSPSENNTDV-----NNALESYKFKFLPEG-TDVATVYSESGSDTLEQSOPK 1213
DB 1698 ESKRLAEQELLETSEVOLLSHQSNTSLNQKKKMSDITQLOTEVEEAQVEQCRNAEKA 1757
QY 1214 KPASTHVGAESNTITTSQ-----NVDDEVDVITVPIFGESEEDYDQGV-VT 1261
DB 1758 KKAITDAAMAEELKEQDTSALHERMKNMEQTIKDL-----QHRLEAEQIALK 1808
QY 1262 GBAVTPSVTDNTLSKIENIEYEVLYLPLAGVYRSKLQLENNVMTNVNKOILNSRFNK 1321
DB 1809 GKKQLQKLEARVRELENELE-----AEQKRAESVKGMRKSERRIKELTYQT 1856
QY 1322 RENFNKVLSDILPDKLTSSNVVVKDPYKFLNKEKDKF---LSSYNYIKDSID 1373
DB 1857 EEDKN-----LMRLQDLVDKQLQKVKAYKQAEAEAEQANTLSKFRKVQHELD 1906

RESULT 45
MYHB_CHKCK
ID MYHB_CHKCK STANDARD; PRT: 1978 AA.

Db 1393 -----ETMEGKKLQRIEISTQOFEEKAASYDKLETKN-----RLQOE 1433
QY 847 LSOLFMEI-YEKEMVNLKLDNDKIKNLEAKVST-----SVKTL 889
Db 1434 LQDVLQDQOLVSNLE--KKQKFDQMLAEKNISSKYADERDRAEAREKETRAL 1491
QY 890 SSSMQPLSITPDQKPE-----VSGANDTSHSTN-----LNN 921
Db 1492 SLARALEBALEAKEELERTNKLKAMEDLVSSKDDVGNVHELESKRTLSQOQVEEMKT 1551
QY 922 SKLFFNILLSGKNKIYQELIGQSSSENFYKILKDSOTFYNESFTNFKSKADDINSL 981
Db 1552 QLEELDELOAAEADAKRLLE-VNMQAMKSOQFERDLQARDEQNEKRRQLLKQHEHETEL 1610
QY 982 NDESK-----RKLEEDINKKLTQLSFDLYNK-----YKLKL 1015
Db 1611 EDERKORALAAAANKKLEVDV-----KLESQVDSANKAREEAIKQRLQOAMKDYORDL 1666
QY 1016 -----ERLF-----DKKKTGVKRMQI-----KLTLLKEQLESKLSNLSN 1050
Db 1667 DDAARAAREEIFATARENEKAKNLEALQLOEDLAAERARKQADLEKEEMAEELASAN 1726
QY 1051 NPKHVLFQSVFNKKAEIAETENTLE-----NTKILKHYKGLVYKYNNGESSPLKTLUS 1106
Db 1727 SGRTSIQD-----EKRLEARIAQLEELDEHSNIETMSDRMKAVQQAQQLNNELATER 1782
QY 1107 ESIQIOTEDNVASLE--NFKVLKSL-----EGKLDNLEKKLKLSYSSGLHLLIAELUEVI 1161
Db 1783 ATAQKNENARQOLERONKELRSKLQMEGAVKSP---KSTIAALEAKTASLEEQLEQBA 1839
QY 1162 KKNKNTGNSPSENNTDVNNAL---ESYKFLPEGTVDATVVSSESGDTLEQSQPKRPAST 1218
Db 1840 REKQAAAKTLRQDKKLKDALQVEDERKQAEYKQDA---EKGNLRLKQ---LKQOLE 1892
QY 1219 HVGAESENTTTS--QNVDDVEDVDVIVPIGESEEDYDGLGOVV 1260
Db 1893 EAEEESQRINANRRKLQRELD-----EATESNDALGREV 1926

RESULT 46
EX5B.BORBU STANDARD: PRT: 1169 AA.
ID EX5B.BORBU
AC 051578;
RX STRAIN=ATCC 35210 / B31;
RA MEDLINE=98065943; Pubmed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Wathley L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi".
RL Nature 390:580-586(1997).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
CC UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
CC STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of

CC ATP) in either 5'- to 3'-or 3'- to 5'-direction to yield 5'-
CC phosphooligonucleotides.
CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVRD SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001164; AAC66981.1; -
CC HSSP: P56255; 1PJR.
CC TIGR: BB0633; -
CC InterPro: IPR004586; RecB.
CC InterPro: IPR000212; UvrD-helicase.
CC Pfam: PF00580; UvrD-helicase; 1.
CC TIGRFAMs: TIGR00609; recB; 1.
CC Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
CC DNA repair; Complete proteome.
CC NP_BIND 18 25 ATP (POTENTIAL).
CC FT SEQUENCE 1169 AA; 137828 MW; B61D63C1C959B91F CRC64;
CC SQ

Query Match 4.0%; Score 338; DB 1; Length 1169;
Best Local Similarity 19.6%; Pred. No. 0.00018;
Matches 288; Conservative 237; Mismatches 455; Indels 490; Gaps 71;

QY 182 NYFDLLRAKLDVNCANDYQIPFNLIKIRANELDVULKLVFGYKRPDLKIKNVGMEDY 241
Db 30 NVVINLIKLYSI-----NEILVLTFT---KKATEEMHTRILKVIEN 69
QY 242 IKKNKKTNIENELIEESK---TIDK-----NKNATKEEKKK----- 277
Db 70 AYSNSKTNEILKEAVEQSKLFSTIFALHALNLFQIETENYSKYKPEKFSKEIDEI 129
QY 278 -----LYQA-----QYDLSYVKNQLEEAHNLISVLEKRIDTLKKNENKELDKIN 323
Db 130 VYDFLRKSDSLIQLADIKDYELKVKFSDAKTEEIVLKIKKAYE---RDTQGLDGLWK 185
QY 324 ETKNPPANGNTPTNLLDKNKKIEEHEKEIKEIAKTIFNIDSLF-----TDPLEEY 378
Db 186 -----TQTAFENILLKKEELIKYNIIEDLRMTDEILSFYKNIHQTKGLEIYS 237
QY 379 LRE-----KKNKIDISAKVETKESTPEYNGVTPLSYDINNALNELNSFGD 428
Db 238 KENDIFKIAETLLKKNFFSTLIEKEFKKSK-----LSPKEL-KINDLICLI 285
QY 429 LINPFDTYKPSKNIYTDNERK-----KFI-NEIKEIK-----IEKKIESDKK 472
Db 286 NIKHEKYKSEDRNKNRNNLKQVILKVEYKILKYIEKELKKTSTNTIDQNYIISNLK 345
QY 473 SY---EDRSKSLNDITPEYKELLNLEYDSKFNNDNLTNPEKMMGR-----YS 518
Db 346 NYLKSDEK-KLINAIAKNRYKIIILIDEAQDLSLIQIEFIIKILTAGIKLFIADPKQIIYS 404
QY 519 YKVEKLTHTHTFASVENSXHNLEKLTALKYMEDYSLRNIVVEKELKYKKNLSIENEI 578
Db 405 FRKADISFN-----KEIKNKINTDAR-IVLKINHRSSKKLIGLPLNKIF 447
QY 579 ETLVENIKKDEQLPEFKITKDNKPEKILEVSDIVKVOVKVLLMNNKIDELKKTQLIL 638
Db 448 NNINNAIADEIE---KIDFTNSLPNQK---NDNNKIVING-----QETEGNIIT 492
QY 639 KNVLEKHNTHVPSNKOENKQEPYIYLIVLKKEIDKLKLVMPKVESLINEKKNIKTEGOS 698
Db 493 TTESEEDIY-----QKTALITIKYLLAYGKTAENNKI-----RNIRKQDITK 533
QY 699 DNSPESTPEITGQATTKPGQAGSALGDSVQAQAEQQAQOQPPVPVPVPEAKAQPPTP 758
Db 534 VILCRGKNEINLIDK-----ALKKEQIQTNKTQEKFL----- 564


```
Qy 759 PAPVNTKNTYVSKLDY-----LEKLYEFNTSYICHKYILVSHSTMNEXKILKOYKITREE 814
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 565 -----KTKESEIFYIKCLDRQSPKTLNVLSSKILNVPWNLQRLIKQDKICLIEE 618
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 815 SKLSCDPLDLFFNTQNNIPWYSWFDSLNLSOLFMEIYEKEMVCHLYKLKNDKI-- 872
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 619 F-----IENIVLEKNEITLINAINKI---TFEKNLWIKIANITDKQKIE 662
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 873 -----KNLEEAKKVSTSVKTL-----SSSQPL-----SLTPQDKPVS 908
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 663 WAKNKINYKGLLIEKGLNLTETYLEIISKYHRENTQSLISTLESILINEEPE-- 720
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 909 ANDTSHSTNLSKLFLFNILSGKNKIYOELIGOKSFNFEYKILKSDTFYNESF- 967
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 721 --EIEKINNNDNESTIE-LMTIHKSKGLGMNIV-----FLNNTPTIENSFF 766
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 968 ---TFVYKSKADDINLSNDESKRKLIEDINKLAKTTLQSLFDLYNKYKILERLFDK 1023
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 767 SKNQFYFYQDG-----KIEYDFPKLEE-----NKKYARLILSEEKNI 806
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1024 ---TVGKYKQWIKKLTLLKEQL--ESKLSLNPKHVLQNFVSF-----FNKKEA 1069
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 807 FVVGATRAKFAFIKINSITSKLEIAKIFTIDIKH---DFNIHEFIGOKRENKKK-- 861
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1070 EIAETENTLNTKILLKHYKGLVYKNGESSPLKTLSEESTQTEDNVASLENFVLSKLE 1129
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 862 ---YNTNVNTKLI-----PPKPIIKNFKKE--YTS----- 887
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1130 GKLXDLNLEKKLSYLSLGLHLLIAELKEVIKKNKNTG--NSPSNTDNNNALESYK 1198
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 888 -----SFSLSLTAQAHH-----KEFYENDFKNIYERKETELDYEPGLE---ET 927
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1189 LPEGTIVATVVSSEGSDFLEQSPKPASTHVGAESENTITTSQNVDDDEVDDVILVPI 1248
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 928 LPKGDIGNIL-----HAAME-----EI-----IFST 949
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1249 SEEDYDDLGQVVTGEAVTPSVIDNLSKIENEYEVLYLKPLAGVYRSKLOLENNVMTFN 1308
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 950 AKDTFDFN-----KKNIEIEIKQIOLNSLNTIEIQ-----NSLAKMI-YNLTYN 996
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1309 VNVKIDILNRNKRNFKNVLESDLIPYKDLTSSNVVVKDYKFLNKE-KDKKFLSSVNY 1367
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 997 IRA---INTRCLDIEELQKEME-----FLIK-----INPEFQOKYL----F 1031
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1368 IKDSIDTIDINFANDVLGYIK-ILSEKYSKDLSDIKKYINDKQGENEKYLPPLNNIETLYK 1426
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1032 DKHEDLHIKLSD---GYLKGIVDLIFKA---NNKIYILD-----YK 1067
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1427 T---VNDKIDLVIHLEAKVLYNYTEKSNVEVKIKELNLYLTKIQDKLADFKKNNFV 1483
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1068 TNYLGNKEDYNTNLENTIKKEYVD---LQYKIYALGIKKILFNKKEY--NQKFGII 1122
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Qy 1484 DLSTDYHNHN---LLTFKFLSTGWFENLAK 1510
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1123 YLFTRAPEDNIECLKSKF-ENGIYF-NLPK 1150
      .  ||:  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 47
MYH6_RAT
ID MYH6_RAT STANDARD; PRT; 1938 AA.
AC P02563; Q63351;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, cardiac muscle alpha isoform (MyHC-alpha).
GN MYH6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=Heart;
RA MEDLINE=90016822; PubMed=2798111;
RX Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Complete nucleotide sequence of full length cDNA for rat alpha
RD cardiac myosin heavy chain.";
RL Nucleic Acids Res. 17:7527-7528(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=90133919; PubMed=2614840;
RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
RD Comparisons suggest a molecular basis for functional differences.";
RL J. Mol. Biol. 210:665-671(1989).
RN [3]
RP SEQUENCE OF 1-167 FROM N.A.
RX MEDLINE=84194059; PubMed=6585819;
RA Mahdavi V., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac alpha- and beta-myosin heavy chain genes are organized in
RD tandem.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:2626-2630(1984).
RN [4]
RP SEQUENCE OF 1512-1938 FROM N.A.
RX MEDLINE=82220036; PubMed=7045682;
RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
RT "Molecular characterization of two myosin heavy chain genes expressed
RD in the adult heart.";
RL Nature 297:659-664(1982).
RN [5]
RP SEQUENCE OF 1872-1938 FROM N.A.
RX STRAIN=Wistar; TISSUE=Heart;
RA MEDLINE=85179510; PubMed=6241892;
RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
RT "Cardiac myosin heavy chain isozymic transitions during development
RD and under pathological conditions are regulated at the level of mRNA
RL availability.";
RL Eur. Heart J. 5:181-191(1984).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15938; CAA34064.1; -
DR EMBL; K01464; AAA41648.1; -
DR EMBL; J00751; AAA41653.1; -
DR EMBL; M32697; AAA41658.1; -
DR PIR; S06005; S06005.
DR PIR; A02988; A02988.
DR HSP; P08799; IWND.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
```


DR InterPro; IPR001609; myosin_head.
DR pfam; PF00063; myosin_head; 1.
DR pfam; PF00612; IQ; 2.
DR pfam; PF01576; Myosin_tail; 1.
DR pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family;
KW Calmodulin-binding.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE
FT DOMAIN 782 811 IQ.
FT DOMAIN 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 656 678
FT DOMAIN 758 772 ACTIN-BINDING.
FT DOMAIN 789 806 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 815 832 CALMODULIN-BINDING (BY SIMILARITY).
FT MOD_RES 128 128 METHYLATION (TR1) (POTENTIAL).
FT MOD_RES 696 696 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 706 706 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 13 13 R -> AP (IN REF. 3).
FT CONFLICT 46 46 V -> A (IN REF. 3).
FT CONFLICT 51 52 VS -> AP (IN REF. 3).
FT CONFLICT 87 87 E -> Q (IN REF. 3).
FT CONFLICT 109 109 MISSING (IN REF. 3).
FT CONFLICT 1566 1566 F -> FF (IN REF. 4).
FT CONFLICT 1575 1575 R -> S (IN REF. 4).
FT CONFLICT 1721 1721 N -> T (IN REF. 4).
FT CONFLICT 1852 1852 T -> N (IN REF. 4).
FT CONFLICT 1870 1870 D -> N (IN REF. 4).
FT CONFLICT 1934 1934 M -> I (IN REF. 4 AND 5).
FT SEQUENCE 1938 AA; 235507 MG; D7BD33FC2B19E3C2 CRC64;
Query Match 4.0%; Score 338; DB 1; Length 1938;
Best Local Similarity 19.6%; Pred. No. 0.00031;
Matches 302; Conservative 262; Mismatches 550; Indels 430; Gaps 71;
QY 28 LVKLEALEDAVLTGYSFOKEKVLNCTSGTAVTTSTPGSKGVASGGSGVASGGS 87
DB 594 LEKNKDPUNETVV---GLYKSSLLK-----MATLFTST-----YASADTGDGSKGKG 638
QY 88 VASGGSVASGGSVASGGSG---NSRRTPNSDSSDAKSYADLKHVRVNYLLTIKELK 143
DB 639 KKKGSFQTVSALHRENKMLTNLRTTHP-----HFVR--CIIPNERK 680
QY 144 YPOLFD--LTNHLMLCDNIHGFYLDIGYEINELLY----- 179
DB 681 APGVMDNPLVMHQLRCNGVLEGRICRKGFP--NRILYGFQRQYRIILNPAAPPEGQFID 738
QY 180 -----KLNFYEDLLRAKLANDVCANDYCOIPN-----L 207
DB 739 SKGAELKGLSUDI-----DHQYKFGHTKVFVKAGLLGLEEMRDERLSRIIT 787
QY 208 KIRANELOVLKLVGYRRKPLD-----NIKDNVG-----KMEDVIK--K 244
DB 788 RIQAARGOLMIRIEF--KKWVRRDALLVIQNIIRAFGVKNPWNPKLYFKIKPLKLSAE 845
QY 245 NKKTIENNELIEESKKTIDKNKATKEEKK--KLYQAQYDLSYNNKQLEEAHNLISYL 302
DB 846 TEKEMANKEEFGRYKDALEKSEARKELEKVMVSLQEKNDLQ---QVQAEQDNLADA 902
QY 303 EKRIIDLTKNE-----NKKELDKTNEKNPPANSNGTNPYLLDKNKKIEHEKEIKEL 357
DB 903 EERCQDLINKIQLEAKYKEMTERLEDEEE---MNA-----ELTAKRKLIED---ECSBL 951
QY 358 AKTIKFNIDSLTDPLELYLLREKNKNIDISAKVE--TKESTEPNPGVTPYPLSYNDI 416
DB 952 KKDI-----DDELTFL-----AKVEKEKHATE-----NKV 976

QY 417 NNALNELNSFGDLINPFDYTKPEPSKNIYTDNERKKFIINEIKKIKIEKKIESDKKSYED 476
DB 977 KNLTEEMAGLDEIIAKL--TKE-----KKALQEAHQ--ALDDLQABEDK----- 1017
QY 477 RSKSLNDITKEYEKLLEIYDSYKFNNDIDLTFEXMMKGRYSYKVE---KLTHHTYFASY 533
DB 1018 ---VNTLTTSKVKLEQQVDD--LEGSLEQEKKVMRDLERAKRKLEGDLKLT--QESIMDL 1070
QY 534 ENSKHNLKLTALKY-----MEDYSLRNIVVEKELKYVKNLISNIENIEIETLVNI 585
DB 1071 ENDKLQLEKLLKKFEFDSIQNSKIEDQALALQOKLKENQAIEELEELE--AERT 1128
QY 586 KKDEQLFEKKITKDENEKPEKILEVSDIVKVQVKVLLMNK--IDELKKKTOLILKNVEL 643
DB 1129 ARAKVEKLRSOLDTRELEETSERLEEAGGATSVQIE---MNKKREAEQKMRDRLEEAFL 1184
QY 644 KNIHVPSYKQENKQEPYLLIVLAKKEDKLKVFMPKPVESLINEEKNKIKTEGSDNSNP 703
DB 1185 QHEA-----TAAALRKKHADSVAELEGQIDNLQ-----RVQKLEKEKSEFKLE--LDDVTS 1234
QY 704 STEGEITGOATTKPGQAGSALLEGDSVQAQAEQKQAOQPPVPVPVPEAKQV---PTPPA 760
DB 1235 HMEQIILKAKANLEKYSR-----TLEQANEYR-----VKLEAQRSLNDDFTQRA 1279
QY 761 PVNNKTEN---VSKLDYLEKLEFNTSYCHKYILVSHSTMNKILKQYKIKTEEESKL 817
DB 1280 KL--QTENGELARQLKEEKEALIIQLTRGKL-----SYTQOMEDLKROL-----BEEGRA 1326
QY 818 SS-----CDPLDLLFNQNNIPVMYSMPFDSLNNLSQLFMEIYKEMVCNLYK 865
DB 1327 KNALAHALQSAHRHCDLRLREQYEEEMEAKAEQLRVLSKANSEVAQWRKYETDAIORTEE 1386
QY 866 LKD--NDKIKNLLEAKKYSTSVKTLSSSSMOPLSLTPQDKPE---VSANDDTSHSTNLNN 921
DB 1387 LEEAKKLAQRLQDAEAEVAVNA--KCSLEKTKHRLQNEIEDLMVDVRSNAAAALDK 1445
QY 922 SUKLENTILSGKNKIYQELIGQKSS-----ENYEKILKSDTFYNES 966
DB 1446 KORNFDKILAEWKOR--YBESQSELESQKARSSTELFKLNKAYEESLEHLETFKREN 1503
QY 967 FTNFVKRADDINSIND-----ESKRKKLEEDINKLKKTL----- 1001
DB 1504 -----KNQOEESLDTQELGEGKKNVHELEKIRKQLEVEKLEQLSALEAEASLEHEEGK 1558
QY 1002 -----QLSFDLYNKKLKER-LDFDKKTVGKYKM-OIKKLTLLKBOLESKLSLNNPKHV 1055
DB 1559 ILRAQLEF---NQIKAEIERKLAEKDEMEQAKRNHLRVVDSLOTSLDAETRSRNEALRV 1615
QY 1056 LONFSVFFNKKEAEIAETENTLNTKILL---KHVGLVKVYNGCESSPLKTLSEESIQ 1111
DB 1616 -----KKMREGDLNEMEIQLSQANRIASBAQKHLK-----NAQAH-----LKDTQLQ 1657
QY 1112 TEDNYASLENPKVLSKLEGLKDNLNLEKKLSYSSGLHLLHIAELKEIKVKNKNTGTNSP 1171
DB 1658 LDDAVRANDD-----LKENAIVERNTLLOAELEELRAVVEQTERSKLAEQEL 1707
QY 1172 SENNTDV-----NNALESYKFL-PECTDVATVVSSEGSTLEQSPKPKPASTHVGAES 1224
DB 1708 IETSERVOLLHQSNTSLNQKKMDADLSQLQTEVEEAVQECRNEAEKAKKAITDAAMMA 1767
QY 1225 NTITTSQ-----NVDDEVDDVIIVITFGSEEDYDGLQV--VTGEAVTFSVIN 1272
DB 1768 BELKREQDTSAHLERMKKNMEQTIKDL-----QHRLDEAEQIALKGGKKLOKLEA 1818
QY 1273 ILSKTENEYEVLYKPLAGVYSKLKOLENNYMTNVNVKDTILNSFNKRENFKNVLES 1332
DB 1819 RYRELENELE-----AEQKRNAESVKGMRKSRERRIKELTYQTEEDKKN----- 1861
QY 1333 LPYKDLTSSNYVVKDPYKFLNKEKRDKF---LSSYNYIKDSID 1373
DB 1862 LVRLQDLVDKLOKLYKAYKQAEAEAEQANTNLKFRKVQHELD 1905

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBERG CELLS
CC OF HODGKIN'S DISEASE.
CC -!- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X64838; CAA46050.1; -
CC DR EMBL: M97501; AAA35693.1; -
CC DR PIR: S22695; S22695.
CC DR Genew: HGNC:10461; RSN.
CC DR MIM: 179838; -
CC DR InterPro: IPR000938; CAP-Gly.
CC DR InterPro: IPR001878; Znf.CCHC.
CC DR Pfam: PF01302; CAP_GLY_2.
CC DR SMART: SM00343; Znf.C2HC; 1.
CC DR PROSITE: PS00845; CAP_GLY_1; 2.
CC DR PROSITE: PS0245; CAP_GLY_2; 2.
CC KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
CC FT DOMAIN 78 120 CAP-GLY 1.
CC FT DOMAIN 143 204 SER-RICH.
CC FT DOMAIN 232 274 CAP-GLY 2.
CC FT DOMAIN 304 331 SER-RICH.
CC FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
CC FT DOMAIN 1408 1421 CCHC-BOX.
CC FT VARSPIC 457 491 MISSING (IN SHORT ISOFORM).
CC FT CONFLICT 1069 1069 D -> E (IN REF. 2).
CC SQ SEQUENCE 1427 AA; 160989 MW; 0A4F166DD94254E8 CRC64;

Query Match 4.0%; Score 336.5; DB:1; Length 1427;
Best Local Similarity 19.3%; Pred. No. 0.00025;
Matches 283; Conservative 249; Mismatches 508; Indels 425; Gaps 60;

QY 43 YSLFQKEKVLNDESGTAVTTSTPG-SKGSVASGGSGSVASGGSGSVASGGSGVA 101
DB 269 YGLFAPVHVKTIGFPST-----TPAKAKANAVRRVMATTSASLKRSPASLSMSVA 323
QY 102 SGGSGSRRTNPDSNDSADKSYAD-----LKHVRN--YLLTIKELKYPQLFDLTN 152
DB 324 SSVSSRPTGLLTETSSRYARKISGTTALQELAKKEQKHIEQLAERDLERAERAKATS 383
QY 153 HM-----LTLCDNIHGFYLDGVEE-INELLYKLNFYDILLRAKLVNDVCANDYQIPF 205
DB 384 HVGEIEQELALAR-----DGHQHVLEAKM---DQLRTMVE---AADREKVEL 427
QY 206 NLKIRANELDVLKLVGFKRPLDNKIDKNVGMEDYIKNNKKTNIENINELIESKKTIDK 265
DB 428 -----LNQLEEKRVVEDLQFVEESTYKGLETTQTKLEHARIKELEOSLLFEKTKADK 482
QY 266 NKNAKTE-----EKKKLYQAYDLISYNKOLEBAHNLISVLEKRIDPLKKNENIKELL 319
DB 483 LORELEDTRVATVSEKSRIMELEKOLALRVQVAE-----LRRLESNKPAGDVDSL 535
QY 320 DKINEIKNPPANGNTPNTLLDNKKTB-----EHEKEIKEIAKTIKFNIDSLFTDPEL 375
DB 536 SLLEIETSS-----LOPKLEVTRTDHQREITS-----LKE 564
QY 376 EYLLREKKNKIDISA-KVETKESTPEPYPNGVTYPLSYNDINNALNELNSFGDLINPD 434
DB 565 HFGARETHQKEIKALYTATEKLSKENE-----SLKSKLEHAN----- 602
QY 435 YTKEPSKNITYDNERRKFFINEKIEKIKIESDKKSYEDR-----SKSLNDITPK 486
DB 603 -----KENSVDIALKSKLETAIAASHQQAAMEELKVSFSKGLGTETA 643
QY 487 EYEKLLNEYDYSKFNNDILTNFCKMMGKRYKYKVEKLTHTHTTFASYENSKHLEKLTKA 546

DB 644 EFAELKTOI-----EKMR-LDYQHEIENLQNOO-----DSERAAHAKEMEALR 685
QY 547 LKYMEDYSLRNIVWELKYKKNLISKIENEIETETIVENIKKDEQ-LPEKKTITKDNKPD 605
DB 686 AKLMK-----VIKEK-----ENLEATRSKLDKAEDOHVEMEDTL--NKLQ 725
QY 606 EKILEVSDI-----VKVQVQVLLMNKIDELKKTOLIKNVELKHNHVPNSYKQENKQBP 661
DB 726 EAEIKVKELEVLQAKCNEQTKVIDNFTSQLKATEKELDLDD----- 766
QY 662 YLIVLKEIDKLVKPMKVESLINEEKKNIKTEGOSDNSEFSTEGEITGQATTPGQQA 721
DB 767 -----ALRKASSGKSEMCKLROOLEAAEKQIKHLETEKNAESSKASITR----- 812
QY 722 GSALGEDSVQAQAEQKQAPVPPVPEAKAQVTPPPAPVNNKTNENSVKL-----DYLEK 777
DB 813 -----ELQGRELK-----LTNLENLSEVSOVKETLEK 840
QY 778 LYEFNTSYICHKYLVS-HSTMNEKILKQYKITREESKLSGCDPLDLFIQNNIPVM 836
DB 841 ELQILKEFAEASEEAVSVORSQOETVKNLHQ--KEEQFNMLSSD-----LEKLENLAD 894
QY 837 YSMFDLNNLSQOLFMEIYEKEMVCLNLYKLDKNDKIKNLEEAQKVSTSVKTLSSSSMOP 896
DB 895 EAKFREKDEREQL-----IKAKELENDIAEIMKMS----- 926
QY 897 LSLTQDRPEVSANDDTSHSTNNLSKLKLFENILSGKNKIYO-ELIGOKSEN--FYE 953
DB 927 -----GDNSSOLTAMNDELRL-----KERDEEQLKUTKANENASFQ 965
QY 954 KILKSDTFYNESFTNFVKSKADDINSLNDESKRKKLEEDINKLKTTLQSLDLNKKYL 1013
DB 966 KSIEDMTVKAESQOEAAK-----KHBEKKELEKRLKLEKMKMESHNOQOELKA 1016
QY 1014 KLERLFDKKKTVGKYMQIKLTLKLEQLESKLNLPKH-VLQNFVFFPKKKEAETA 1072
DB 1017 RYERATSETKTKEETILQNLQTL--DTEDEKLKGAARENSGLLOELE---ELRQADKA 1071
QY 1073 ETEPLENTKILLKHYKGLVYKNGESSPLKTLSESIQEDNYASLENFKVLKLEGLK 1132
DB 1072 KAAQTAEDAMQINEQ-----WTEKETET---LASLEDTK---QTNAKL 1108
QY 1133 KDNLEKKKLSYLSGLHLIAELKEVIKNKNYTGNSPENNTDYNNALESYKFLPEG 1192
DB 1109 QNELDTLKE-----NNLKNVEELNKS-----ELLTVENOKMEEFKEIETL 1150
QY 1193 TDVAIVWSESGSDTLEQSQPKPASTHVGAESNTTITSQNVDE---VDDVII-----V 1243
DB 1151 KQAAAKSQOLS-ALQOENVK--LABELGRSDEVTSHOKLEERSVLNQLEMKKRES 1207
QY 1244 PIFGESEEDYDGLGVVTGEAVTPSVIDNLSKIENEYEVL-----YLKPLAGVYRSL-- 1296
DB 1208 KFIKDADEKASLQKSITSALLTEKDAELEKLRNEVTVLGENASAKLSHVQVTL 1267
QY 1297 -KKOLENNVMTFNVAVKOILNSRFNKRNFKNVLESDLIPYKDLTSSNVVVKDPYKFLNK 1355
DB 1268 DKVKLEKLVKNELELQKE-----NKRQLSSSGNTDTQADEDERA-----Q 1308
QY 1356 EKRDFLSINYIKOSIDTIDINFAN-DVLGYKYLISE-KYKSDLSIKKIYKNDKQENK 1413
DB 1309 ESQIDFL-----NSVIVDQKQKQDLKMKYEMMSEALNGDGLNNYSDQKQSK 1361
QY 1414 YLPFLNNIETLYKTVNDKIDLEVIH 1438
DB 1362 KKP-----RLPDCIDCDFDLH 1377

RESULT 51
DMD_HUMAN STANDARD; (PRT; 3685 AA.
AC P11532; Q14169; Q14170;
DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN Dystrophin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88194521; PubMed=3282674;
RA Koenig M., Monaco A.P., Kunkel L.M.;
RT "The complete sequence of dystrophin predicts a rod-shaped
cytoskeletal protein.";
RL Cell 53:219-228(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345106; PubMed=2668885;
RA Rosenthal A., Speer A., Billowitz H., Cross G.S., Forrest S.N.,
RA Davies K.E.;
RT "Two human cDNA molecules coding for the Duchenne muscular dystrophy
(DMD) locus are highly homologous.";
RL Nucleic Acids Res. 17:5391-5391(1989).
RN [3]
RP SEQUENCE OF 1-497 FROM N.A.
RX MEDLINE=87273512; PubMed=3607877;
RA Koenig M., Hoffman E.P., Bertelson C.J., Monaco A.P., Feener C.,
RA Kunkel L.M.;
RT "Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and
preliminary genomic organization of the DMD gene in normal and
affected individuals.";
RL Cell 50:509-517(1987).
RN [4]
RP SEQUENCE OF 404-1137 FROM N.A.
RX MEDLINE=8811512; PubMed=3428261;
RA Cross G.S., Speer A., Rosenthal A., Forrest S.M., Smith T.J.,
RA Edwards Y., Flint T., Hill D., Davies K.E.;
RT "Deletions of fetal and adult muscle cDNA in Duchenne and Becker
muscular dystrophy patients.";
RL EMBO J. 6:3277-3283(1987).
RN [5]
RP SEQUENCE OF 665-722; 2098-2204 AND 2305-2366 FROM N.A.
RX MEDLINE=89083552; PubMed=3205741;
RA Chamberlain J.S., Gibbs R.A., Ranier J.A., Nguyen P.N.,
RA Caskey C.T.;
RT "Deletion screening of the Duchenne muscular dystrophy locus via
multiplex DNA amplification.";
RL Nucleic Acids Res. 16:11141-11156(1988).
RN [6]
RP SEQUENCE OF 2147-2204 FROM N.A.
RX MEDLINE=89345155; PubMed=2569720;
RA Blondin L.A.J., den Dunnen J.T., van Paassen H.M.B.,
RA Wapenaar M.C., Grootscholten P.M., Ginjaar H.B., Bakker E.,
RA Pearson P.L., van Ommen G.J.B.;
RT "High resolution deletion breakpoint mapping in the DMD gene by whole
cosmid hybridization.";
RL Nucleic Acids Res. 17:5611-5621(1989).
RN [7]
RP SEQUENCE OF 2305-2364 FROM N.A.
RX Huth A., Will K., Speer A., Bauer D.;
RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [8]
RP REVIEW ON DMD POINT MUTATION VARIANTS.
RX MEDLINE=95038763; PubMed=7951253;
RA Roberts R.G., Gardner R.J., Bobrow M.;
RT "Searching for the 1 in 2,400,000: a review of dystrophin gene point
mutations.";
RL Hum. Mutat. 4:1-11(1994).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=94320940; PubMed=8045556;
RA Rininsland F., Reiss J.;
RT "Microlesions and polymorphisms in the Duchenne/Becker muscular

dystrophy gene.";
RL Hum. Genet. 94:111-116(1994).
RN [10]
RP VARIANT DMD ARG-54
RX MEDLINE=94004962; PubMed=8401582;
RA Prior T.W., Papp A.C., Snyder P.J., Burghes A.H.M., Bartolo C.,
RA Sedra M.S., Western L.M., Mendell J.R.;
RT "A missense mutation in the dystrophin gene in a Duchenne muscular
dystrophy patient.";
RL Nat. Genet. 4:357-360(1993).
RN [11]
RP VARIANTS DMD HIS-365; TRP-2191 AND ARG-2937.
RX MEDLINE=95152525; PubMed=7849724;
RA Nigro V., Nigro G., Esposito M.G., Comi L.I., Molinari A.M.,
RA Puca G.A., Politano L.;
RT "Novel small mutations along the DMD/BMD gene associated with
different phenotypes.";
RL Hum. Mol. Genet. 3:1907-1908(1994).
RN [12]
RP ALTERNATIVE SPLICING (DYSTROPHIN-1 AND -2).
RC TISSUE=Retina;
RA White R.A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [13]
RP ALTERNATIVE SPLICING (DYSTROPHIN-3).
RC TISSUE=Brain;
RX MEDLINE=89181947; PubMed=2648158;
RA Feener C.A., Koenig M., Kunkel L.M.;
RT "Alternative splicing of human dystrophin mRNA generates isoforms at
the carboxy terminus.";
RL Nature 338:509-511(1989).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
PLASMA MEMBRANE.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY
ALTERNATIVE SPLICING; THE 3 SHORTER VARIANTS ARE CALLED
DYSTROPHIN-1, -2 AND -3.
CC -!- DISEASE: Defects in DMD are the cause of Duchenne muscular
dystrophy (DMD) and Becker muscular dystrophy; a sex-linked recessive
most common form of muscular dystrophy; a sex-linked recessive
disorder. It typically presents in boys aged 3 to 7 year as
proximal muscle weakness causing waddling gait, toe-walking,
lordosis, frequent falls, and difficulty in standing up and
climbing up stairs. The pelvic girdle is affected first, then the
shoulder girdle. Progression is steady and most patients are
confined to a wheelchair by age of 10 or 12. Flexion contractures
and scoliosis ultimately occur. About 50% of patients have a lower
IQ than their genetic expectations would suggest. There is no
treatment. BMD resembles DMD in hereditary and clinical features
but is later in onset and more benign.
CC -!- DISEASE: Defects in DMD are a cause of X-linked dilated
cardiomyopathy (XLCM).
CC -!- MISCELLANEOUS: THE DMD GENE IS THE LARGEST KNOWN GENE IN HUMANS.
IT IS 2.4 MILLION BASE-PAIRS IN SIZE AND COMPRISES 79 EXONS.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -!- DATABASE: NAME=DMD; NOTE=Dystrophin Mutation Database;
WWW="http://www.dmd.nl/database.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06179; CAA29545.1; ALT_SEQ.
DR EMBL; X06178; CAA29544.1; -.

Db 2427 LAPGLTTIGASTQVTLVTPVTKETAISKLEMPSSLMLEVPALADFNRAWTELTDL 2486
QY 1256 --LGQVTTGEAVTPSIDNLSKIENEYEVL-----YKPLAGVYRSKKQLNNVM 1305
Db 2487 SLLDQVIKRSQVWVGDLDEINEMIKQKATQDLEQRPQLEELITAAQNLKNK----- 2540
QY 1306 TFWNVKQILNRSFKNREKFNKVDLSIPYKDLTSSNVVVKDPYKFLN-KEKDKFLSS 1364
Db 2541 TSQNEARTIITDRIERIQNDQVQHLQNRQ--QLNEMKLDSTQWLKEAKEAQVLGQ 2598
QY 1365 YN-----YIKDSI-----DTDINFAND-----VLGYIK----- 1387
Db 2599 ARAKLESWEGYTVDAQKKTETTKQAKLQKQWQTVNDVANDLALLRLDYSADDTK 2658
QY 1388 --ILSEKYSKSDLSIKKYINDKQ-----ENKYLPLNNIETLYKTVDK 1431
Db 2659 VHMITENINASWRSIHKVRSEREALEETHRLLOQFPIDLEKFLAWLTAETANVLQD- 2717
QY 1432 IDLFVHLEAKVLNTEKSNVEVKEIKELNYLKTQDKLADPKNNFVGIADLSTYNNH 1491
Db 2718 -----ATKRELEDSKGVKELM--KQWDLQGETEATHTDVVH 2753
QY 1492 NNLLTKFLSTGMVFENLAKTVLS-----NLLDGNLQGM-----LNISQH 1530
Db 2754 N-----LDENSKILRSLEGSDDAVLLQRRLLDNNMKNKWSLLRKKSLNIRSH 2799

RESULT 52
DESP_HUMAN
ID DESP_HUMAN STANDARD: PRT: 2871 AA.
AC P15924; Q14189; Q75993; Q9UHN4;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen).
GN DSP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RC TISSUE=ForeSkin;
RX MEDLINE=92115697; PubMed=1731325;
RA Virata M.L.A., Wagner R.M., Parry D.A.D., Green K.J.;
RT "Molecular structure of the human desmoplakin I and II amino
RT terminus";
RL Proc. Natl. Acad. Sci. U.S.A. 89:544-548(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM DPI).
RA Phillips S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1120-2871 FROM N.A. (ISOFORM DPI).
RC TISSUE=ForeSkin;
RX MEDLINE=90153880; PubMed=1689290;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RT "Structure of the human desmoplakins. Implications for function in
RT the desmosomal plaque";
RL J. Biol. Chem. 265:2603-2612(1990).
RN [4]
RP ERRATUM.
RX MEDLINE=90361712; PubMed=2391353;
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,
RA Angst B.D., Nilles L.A.;
RL J. Biol. Chem. 265:11406-11407(1990).
RN [5]
RP SEQUENCE OF 2854-2871 FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=20062965; PubMed=10594734;
RA Whittcock N.V., Ashton G.H., Dopping-Hepenstal P.J., Gratian M.J.,
RA Keane F.M., Eady R.A.J., McGrath J.A.;

*Striate palmoplantar keratoderma resulting from desmoplakin haploinsufficiency";
J. Invest. Dermatol. 113:940-946(1999).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=98012209; PubMed=9348293;
RA Kowalczyk A.P., Bornslaeger E.A., Borgwardt J.E., Palka H.L.,
RA Dhaliwal A.S., Corcoran C.M., Denning M.F., Green K.J.;
RT "The amino-terminal domain of desmoplakin binds to plakoglobin and
RT clusters desmosomal cadherin-plakoglobin complexes";
J. Cell Biol. 139:773-784(1997).
CC !- FUNCTION: MAJOR HIGH MOLECULAR WEIGHT PROTEIN OF DESMOSOMES.
CC INVOLVED IN THE ORGANIZATION OF THE DESMOSOMAL CADHERIN-
CC PLAKOGLOBIN COMPLEXES INTO DISCRETE PLASMA MEMBRANE DOMAINS AND IN
CC THE ANCHORING OF INTERMEDIATE FILAMENTS TO THE DESMOSOMES.
CC !- SUBUNIT: HOMODIMER.
CC !- SUBCELLULAR LOCATION: INNERMOST PORTION OF THE DESMOSOMAL PLAQUE.
CC !- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DPI/DPI1 (SHOWN HERE) AND
CC DPI1/DPI2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC !- TISSUE SPECIFICITY: DPI IS APPARENTLY AN OBLIGATE CONSTITUENT OF
CC ALL DESMOSOMES; DPI1 RESIDE PREDOMINANTLY IN TISSUES AND CELLS OF
CC STRATIFIED ORIGIN.
CC !- DOMAIN: THE N-TERMINAL REGION IS REQUIRED FOR LOCALIZATION TO THE
CC DESMOSOMAL PLAQUE AND INTERACTS WITH THE N-TERMINAL REGION OF
CC PLAKOPHILIN 1. THE C-TERMINAL REGION INTERACTS WITH INTERMEDIATE
CC FILAMENTS.
CC !- DISEASE: DEFECTS IN DSP ARE A CAUSE OF STRIATE PALMOPLANTAR
CC KERATODERMA II (PPK2, KPP2 OR SPPK2), CHARACTERIZED BY SKIN
CC THICKENING IN THE PALMS (LINEAR PATTERN) AND THE SOLES (ISLAND-
CC LIKE PATTERN) AND FLEXOR ASPECT OF THE FINGERS; AND RARELY BY
CC ABNORMALITIES OF THE NAILS, THE TEETH AND THE HAIR.
CC !- SIMILARITY: CONTAINS 17 PLECTIN REPEATS.
CC !- SIMILARITY: CONTAINS 2 SPECTRIN REPEATS.
CC !- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M77830; AAA85135.1; -;
CC EMBL; AL031058; CAA19927.1; -;
CC EMBL; J05211; AAA35766.1; -;
CC EMBL; AF139065; AAF19785.1; -;
CC PIR; A35536; A35536.
CC Genew; HGNC:3052; DSP.
CC MIM; 125647; -;
CC InterPro: IPR001101; Plectin.repeat.
CC InterPro: IPR002017; Spectrin.
CC Pfam; PF00681; Plectin; 10.
CC SMART; SM00250; PLEC; 17.
CC SMART; SM00150; SPC; 1.
CC Repeat; Coiled coil; Phosphorylation; Cytoskeleton;
CC Structural protein; Alternative splicing.
CC DOMAIN 1 1056 GLOBULAR 1.
CC DOMAIN 1057 1945 CENTRAL FIBROUS ROD DOMAIN.
CC DOMAIN 1946 2871 GLOBULAR 2.
CC REPEAT 347 447 SPECTRIN 1.
CC REPEAT 858 952 SPECTRIN 2.
CC DOMAIN 1018 1945 COILED COIL (POTENTIAL).
CC REPEAT 2009 2045 PLECTIN 1.
CC REPEAT 2046 2083 PLECTIN 2.
CC REPEAT 2084 2121 PLECTIN 3.
CC REPEAT 2122 2159 PLECTIN 4.
CC REPEAT 2163 2197 PLECTIN 5.
CC REPEAT 2198 2233 PLECTIN 6.
CC REPEAT 2251 2288 PLECTIN 7.
CC REPEAT 2289 2326 PLECTIN 8.
CC REPEAT 2327 2364 PLECTIN 9.
CC REPEAT 2365 2402 PLECTIN 10.

RC STRAIN-Canton-S; TISSUE-Pupae;
 RX MEDLINE=89384556; PubMed=2506434;
 RA George E.L., Ober M.B., Emerson C.P. Jr.;
 RT "Functional domains of the Drosophila melanogaster muscle myosin
 RL heavy-chain gene are encoded by alternatively spliced exons.";
 RN Moll. Cell. Biol. 9:2957-2974(1989).
 RP [2]
 RQ SEQUENCE OF 1-264 FROM N.A.
 RX MEDLINE=87280141; PubMed=3038896;
 RA Wassenberg D.R. II, Kronert W.A., O'Donnell P.T., Bernstein S.I.;
 RT "Analysis of the 5' end of the Drosophila muscle myosin heavy chain
 RL gene. Alternatively spliced transcripts initiate at a single site and
 RN intron locations are conserved compared to myosin genes of other
 RP organisms.";
 RQ J. Biol. Chem. 262:10741-10747(1987).
 RP [3]
 RQ SEQUENCE OF 333-614 FROM N.A.
 RX STRAIN-Canton-S; TISSUE-Embryonic muscle;
 RA MEDLINE=91330870; PubMed=1907912;
 RT Kronert W.A., Edwards K.A., Roche E.S., Wells L., Bernstein S.I.;
 RL "Muscle-specific accumulation of Drosophila myosin heavy chains: a
 RN splicing mutation in an alternative exon results in an isoform
 RP substitution.";
 RQ EMBO J. 10:2479-2488(1991).
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- ALTERNATIVE PRODUCTS: At least 15 isoforms; AAAAA (shown here),
 CC BBBBB, BABBB, 3b, 7b, 7c, 9b, 9c, 11b, 11c, 11d, 11e, 15b and
 CC 18; may be produced by alternative splicing. Exons 3, 7, 9, 11 and
 CC 15 are mutually exclusive splicing exons and exon 18 is included
 CC or excluded.
 CC -!- TISSUE SPECIFICITY: Expressed in larval and adult muscles.
 CC Isoforms containing exon 9a are expressed in indirect flight
 CC muscles, exons 9a and 9b are expressed in jump muscles, exons 9b
 CC and 9c are expressed in other larval and adult muscles.
 CC -!- DOMAIN: Alternative splicing exons contribute to the specialized
 CC contractile activities of different muscle types. Exon 3 encodes
 CC the hydrophobic pocket adjacent to the ATP-binding site, exon 9 is
 CC adjacent to the actin-binding domain, exon 11 is involved in
 CC actin-binding, exon 15 in the S2 hinge and exons 18 and 19 the
 CC non-coiled tail region.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M61229; AAA28686.1; -;
 DR EMBL; M61229; AAA28687.1; -;
 DR EMBL; J02788; AAA28706.1; -;
 DR EMBL; J02788; AAA28707.1; -;
 DR EMBL; X60196; CAA42752.1; -;
 DR EMBL; X60196; CAA42753.1; -;
 DR EMBL; X60196; CAA42754.1; -;
 DR PIR; A28492; A28492.
 DR PIR; A32491; A32491.
 DR PIR; B32491; B32491.
 DR HSP; P08799; 1MND.
 DR FlyBase; FBgn0002741; Mhc.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR004009; Myosin_N.

DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 7.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 4.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MYSC; 1.
 DR PROSITE; PS00096; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Calmodulin-binding; Alternative splicing;
 KW Multigene family.
 FT DOMAIN 1 779 MYOSIN HEAD-LIKE.
 FT DOMAIN 780 809 IQ.
 FT DOMAIN 802 1927 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (BY SIMILARITY).
 FT VARSPLIC 69 116 YRAKLY -> TRDLKKLLQVNPVKYAEADMSNLTYLN
 FT DASVHLNLRORYNKLIV (IN ISOFORM 3B).
 FT ISOFORM BBBBA AND ISOFORM BABDB).
 FT DICTLNTNIDYHIVSQGKVTVASIDDAEESLTD -> EY
 FT CLLSNNIYDRIYVRSQGGKTTPSVNDGGEWAVID (IN
 FT ISOFORM 7B).
 FT VARSPLIC 298 332 DICTLNTNIDYHIVSQGKVTVASIDDAEESLTD -> EM
 FT VFGHGHGIDYGPICQGRTRIPGVNDGEFELTD (IN
 FT ISOFORM 7C).
 FT VARSPLIC 298 332 DICTLNTNIDYHIVSQGKVTVASIDDAEESLTD -> EM
 FT CFLSDNIYDYNVSGKVTVPNMDGGEFQIAD (IN
 FT ISOFORM 7D AND ISOFORM BBBBA).
 FT VARSPLIC 469 525 YNGPEQLCINFTNEKLOQFNHIMFVMEQEEYKKEGINWDF
 FT IDFGMDLLACIDLIEK -> YNGPEQLCINFTNEKLOQFN
 FT HHMEVLEQEEYKKEGIDWAFIDFGMDLLACIDLIEK (IN
 FT ISOFORM 9B, ISOFORM BBBBA AND ISOFORM
 FT BABDB).
 FT VARSPLIC 469 525 YNGPEQLCINFTNEKLOQFNHIMFVMEQEEYKKEGINWDF
 FT IDFGMDLLACIDLIEK -> YNGPEQLCINFTNEKLOQFN
 FT HHMEVLEQEEYKKEGIDWAFIDFGMDLLACIDLIEK (IN
 FT ISOFORM 9C).
 FT VARSPLIC 723 761 YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT -> YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT K (IN ISOFORM 11B AND ISOFORM BBBBA).
 FT VARSPLIC 723 761 YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT -> YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT K (IN ISOFORM 11C).
 FT VARSPLIC 723 761 YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT -> YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT K (IN ISOFORM 11D AND ISOFORM BABDB).
 FT VARSPLIC 723 761 YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT -> YQILNPRGKIDLCPPKASKVLIESTELNEDLYRLGHTK
 FT -> YMLAPAIAMAAEKVAKNAAGKCLEAVGLDPPDWRIGT
 FT K (IN ISOFORM 11E).
 FT VARSPLIC 1216 1241 AEHRQTCNHELNTNTRACDQLGRDK -> AEKEKNEYGQ
 FT LNDURAGVDHITNEK (IN ISOFORM 15B AND
 FT ISOFORM BABDB).
 FT VARSPLIC 1936 1936 P -> I (IN ISOFORM 18).
 FT VARSPLIC 1937 1962 MISSING (IN ISOFORM 18).
 FT CONFLICT 43 44 EK -> RE (IN REF. 2).
 FT CONFLICT 68 68 E -> K (IN REF. 2).
 FT CONFLICT 215 215 L -> M (IN REF. 2).
 SQ SEQUENCE 1962 AA; 224480 MW; F5A888932E414F7F CRC64;
 Query Match 4.0%; Score 333.5; DB 1; Length 1962;
 Best Local Similarity 18.3%; Pred. No. 0.00044;
 Matches 304; Conservative 284; Mismatches 577; Indels 493; Gaps 74;
 QY 19 CVTHESYOELVKLEALEDAVLITGYSLFQEKMKVNLNCTSGTAVTSTPTPGSKGVSAGSGS 78
 Db 585 CVSYNITGWEKKNKDLNDLVVDQFKKSNKLLI-----EIFADHAGQSG----- 629
 QY 79 GGSVASGGSVASGGSVASGGSVASGGSVASGGSVASGGSVASGGSVASGGSVASGGSVASGGS 138
 Db 630 -----GGEQAKGGRGKGGGFATVSSAYKEQLNSLMTTLRSTQ-----PHFVR--CII 675

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U75316; AAB37320.1; -
 CC HSP; P08799; IMND.
 CC InterPro; IPR000048; IQ_region.
 CC InterPro; IPR004009; Myosin_N.
 CC InterPro; IPR002928; Myosin_tail.
 CC InterPro; IPR001609; myosin_head.
 CC Pfam; PF00063; myosin_head; 1.
 CC Pfam; PF00612; IQ; 2.
 CC Pfam; PF01576; Myosin_tail; 1.
 CC Pfam; PF02736; Myosin_N; 1.
 CC PRINTS; PR00193; MYOSINHEAVY.
 CC ProDom; PD000355; myosin_head; 1.
 CC SMART; SM00015; IQ; 1.
 CC SMART; SM00242; MYSC; 1.
 CC PROSITE; PS00096; IQ; 1.
 CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 KW Calmodulin-binding.
 FT DOMAIN 1 780 MYOSIN HEAD-LIKE.
 FT DOMAIN 781 810 IQ.
 FT DOMAIN 839 1935 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT DOMAIN 655 677 ACTIN-BINDING.
 FT DOMAIN 757 771 ACTIN-BINDING.
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 695 695 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1935 AA; 223110 MW; 149CDBFD910DBB08 CRC64;
 Query Match 3.98; Score 332.5; DB 1; Length 1935;
 Best Local Similarity 18.78; Pred. No. 0.00047;
 Matches 286; Conservative 266; Mismatches 576; Indels 401; Gaps 66;
 QY 28 LVKKLEADAVLTGYSLFQKRMVLMNEGTSQTAVTSTPGSKGVSASGGSGSVASGGS 87
 DB 594 LQNKDPLNETV--DLYKSKLLSLNLFANYAGADTPVEKCK-----GK 637
 QY 88 VASGGSVASGGSVASGSG-----NSRTPNSDSDSDAKSYADLKHVRNLYLTIKELK 143
 DB 638 AKGSGSTVSALHRENKLNKMLNLTNRSTP-----HFVR--CIIPNETK 679
 QY 144 YPOLFD--LTHNMLTCDNTHGFYLDGYEEINELLYK-----LN-----FYFD 186
 DB 680 SPGVIDNPLVMHQRCNGVLEGIRCRKGF--NRILYGFQRYRILNPAAPGQFID 737
 QY 187 LLRA--KLNDVCANDYQIQIPN-----LKRANELDYLK 218
 DB 738 SRKGAELGLSLDIDHNOYKFGHTKVFVKAGLLGLEEMRDERLSRITRIQAOSRGVLS 797
 QY 219 KLVFGYKKPLD-----NIK-----DNVGMEDYK-----KNKKTIENTNEL 255
 DB 798 RMEF--KKLLERDLSLTIQNIARFMSVKNPWNKLYFKIKPLLESAAETEKEMATKKEE 855
 QY 256 IEESKKTIDKNKATKEEK--KLYQAQYDLSYINQLEBAHNLISVLEKRIDTLAKNE 313
 DB 856 FGRLEALESEARKELEKMWLSLQEKNDLQ---OVQAEQNDLSDSEERCOLDLKNK 912
 QY 314 -----NIKELDKINEIKNPPANGTPTNLLDKNKKIEHEKEIKETIKFNIDSL 368
 DB 913 IOLEAKYKEMTERLEDEE-----EMNAELTAKKNVEDECSLARDID-- 955
 QY 369 FTDPLEYLYREKKNIDISAKVE--TKESTEPNEYPNGVTPLSYNDINNALNELNPSG 427
 DB 956 -----DLELT-----AKVEKEKHATE-----NKVKNLTEEMAGLD 986
 QY 428 DLINPFDTYTPSKNIYTDNRKKFINEIKIEKKIESDKKSYEDSKSLNDITKE 487
 DB 987 EIIAKL--TKE-----KKALQEAHQ--ALDDLOAEEDK-----VNTLTKA 1023

QY 488 YEKLLNEIYDSKFNNDLTFNFKMMGRYSYKVE---KLTHHTNFASYENSKHNLEKLT 544
 DB 1024 KVKLEQHVDD--LAGSLBQEKVMDLERAKRKLEGDKLKT--QBSIMDLENDKQDLDERL 1080
 QY 545 KALKY-----MEDYSLRNIVVEKELKYKNLSKIENIETLVENIKKDEEQLFKK 596
 DB 1081 KKKDFELNALNARTEDEQALGSLQKKLQELQARIEELEE---AERTAKAVEKLRS 1138
 QY 597 ITKDENKPEKILEVSDIVKQVQKVLNKK--DELEKKTQLILKNVELKHNHVPNSYK 654
 DB 1139 LSRELEISERLEEAGGATSVQIE-----MNRKRAEFQKMRDRLEAEATLQHEA----TAA 1190
 QY 655 QENKQEPYLLIVLKKEDIKLVFMPKVBESLNEEKKNKTTEGQSDNSPSPSEGETGOAT 714
 DB 1191 ALRKHADSVAEELGEQIDNLQ-----RVKQLEKSEKFELE--LDDVTSNMQIILAKAN 1244
 QY 715 -----TKPQOAGSALBGDSVQAQAEQQAQPPVPVPEAKAQVPTPPAPVNNKTEN 768
 DB 1245 LEKMCRTLEDQMNHRSKAETQRSVNDLTQRAKLQOTENGELSRLQDEKALISQLTR- 1303
 QY 769 VSKLDYLEKL-----YEFNTSYICHKYLIVSHSTMNEKILKQYKITTEESKLSSC 820
 DB 1304 -GKUTYTOOLEDLKQLEEEVKAKNALAHALOSARHAA--DULREQYE--EETETKAE-- 1356
 QY 821 DPLDLLENQNNIPVMYSMFDLSNLSQLEMEIYKEMVNCNLYKLD--NDKIKNLLBEA 879
 DB 1357 -----LQRVLSKANSEVAQ--WRTYETDAIQORTELEEAKKKLAQLQDA 1400
 QY 880 KKVSTSVKTLSSSMOPLSLTPQDKPE---VSANDDTSHSNLNNLSKLFENILSLGNK 936
 DB 1401 EEAVEAVNA--KCSSLEKTKHRLQNEIEDLVMDVDSRNSAAAAALDKKQNFOKILAEWK 1459
 QY 937 NIYELIGOKSS-----ENFEYKILKSDTFYNESFTNFVSKADDDINSL 981
 DB 1460 --EESOSELESSOKEARSSTELFKLKNAYEESLEHLETSKREN-----KNQJESOL 1512
 QY 982 ND-----ESRKKLEEDINKLTKLT-----OLSFOLYNKYK 1012
 DB 1513 TEQLGSSGKTHLEKVRKQLEAEKLEQSALEESSEASLEHEEGKILRAQLEF---NQIK 1569
 QY 1013 LKLER-LFDKKKTVGKYKM-QIKKLTLLKEQESKLNNPKHVLFQNFVFNKKKAE 1070
 DB 1570 AMERKLAEKDEEMEQAKRNHLRVVDSLSQTSIDAETRNEALRV-----KKMEGD 1621
 QY 1071 IAEFTENTLENTKILLKHVKGIVKYNGESSPLKTLSESIQIOTEDNYASLENFKVLSKLEG 1130
 DB 1622 LNEIQLSHANRMAAAEQKVK-----SLQSLKQTOQLDDAVRAND-- 1666
 QY 1131 KLDNLEKLLKSLYSSGLHHLIAELKEVINKKNYTGNSPSENNTDV-----NNALE 1183
 DB 1667 -LKENIAIVERNNLLQAELEELRAVVBQTSRSLAEQELIETSERVQLLHSQNTSLIN 1725
 QY 1184 SYKKFLPBGTDVATVWSESGDTLEQSQPKKPASTHVGAESENTTTSQ-----N 1232
 DB 1726 QKKMEADLSOLOTEVEEAQVQSRNAEAKKAITDAAMMAEELKKEQDTSALHERMKKN 1785
 QY 1233 VDDEVDVVIIPFGESEEDVDLGOV--VTGEAVTPPSVIDNLSKIEREYEVLYLKPLAG 1291
 DB 1786 MEQTKDL-----QHRUDEAEQALKGGKQKQKLEARVRELENELE----- 1827
 QY 1292 VYRSLKQLENNVMTFNVNVDILNSRFNKNFNKLVLE--SDLPYKDLTSSNYVVKD 1350
 DB 1828 ---AEQKRNAEVSGMKRSERRIKELTYQTEEDRKNLLRLQDLVDKQLKVAY----- 1878
 QY 1351 KFLNK-----EKDKFLSSSYNYIKDSID 1373
 DB 1879 ---NRQAEAEQANTLSKPKRVQHELD 1904

RESULT 55
 SCPI_HUMAN
 ID SCPI_HUMAN STANDARD; PR; 976 AA.

Db 694 -----KLOKEIDKRCQH-----KIAEMVALMEKHHQYDKIEERDSE----- 731
QY 956 LKGDSTFVNESTFNKSKADDINSNDKSRKKLEEDINKLKTQLSFDLYNKYKLL 1015
Db 732 -----IGLYKSKEQSSSL-----RASLEIELSNLKAEL-----LSVKKQLEI 769
QY 1016 ERLFDKKTGKYKMQIKKLT-LKKEQESKLSLNNPKHVQLNFVSFFNKKKEAEIAET 1074
Db 770 ER--EKEE--KLKREAKENTATLKEKKDKK-----TQTF----- 799
QY 1075 ENTLENTKILLKHVKGLVKNYNGESSPLKTLSESIQTEDNYASLENPKVLSKLEGLKD 1134
Db 800 --LLETPEIYWK-----LDSKAVPSQTVSR-----NFTSVDH----- 829
QY 1135 NLNLEKKLSYL-SSGLHLLIAELKEVKNKNYTGNSPS-----ENNTDVNNALESYKK-- 1187
Db 830 --GISKDKRDYLWTSAKNTLSTPLP-----KAYTVKTPKPKLOQRENLIPIESKKR 882
QY 1188 -----FLPGGTQVATVVSSESGDTLEQ-SQPKKPASTHVGAES-----NTTIT 1229
Db 883 KMAFEFDINDSSSETTDLMSVSE--BETLKTLYRNNPPASHCLCVKTPKKAPSLAT 938

RESULT 56
RASO_SULSO STANDARD; PRT; 864 AA.
AC Q97WH0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 AtPase.
GN RAD50 OR SSO2249.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Cafalonieri F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE006829; AAK42417.1; .
DR InterPro; IPR003439; ABC transportr.
DR InterPro; IPR003405; SMC.C.
DR InterPro; IPR003395; SMC.N.
DR Pfam; PF02463; SMC.N; 1.
DR Pfam; PF02483; SMC.C; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.

FT NP_BIND 30 37 : ATP (BY SIMILARITY).
FT DOMAIN 164 701 COILED COIL (POTENTIAL).
SQ SEQUENCE 864 AA; 101601 MW; 657076AEA9B709FC CRC64;

Query Match 3.9%; Score 331.5; DB 1; Length 864;
Best Local Similarity 20.8%; Pred. No. 0.00021;
Matches 230; Conservative 200; Mismatches 321; Indels 357; Gaps 55;

QY 192 LNDVCANDYCOIPENLKIR-----ANELDVLKLVFGY-----RKPLDNI--KDNVG 236
Db 8 LTNFSLSEHSIQPMGEINLVINGONGACKSSIIDGIVFSLFRHSRGNNDNLIRKGSNRG 67
QY 237 KMEDYIRKKNKTIENINELIEESKKTIDKN-----KNAT-KEEEKKKLYQAYDLSIYNK 290
Db 68 SVTLYSNEKDKIEIRDIRSTTDEDLRNQFPPIARSATVVSNEIEKILGIDKDIAL-- 124
QY 291 QLEAHNLISVLEKRIDTLKKNENIKELLDKINEIKNPPANSNGTPTNLLDKNKKIEEH 350
Db 125 -----STIIVRQGEID--KILENFQETMGKI-----LKLEI 154
QY 351 EKEIKEIATAKTIKFNIDSLFTDPLEYVLRKKNKIDISAKVETKESTPEPNPVGTYP 410
Db 155 EKLDSRGPIVEF-----RKNLE----- 172
QY 411 LSYNDINNALNELNSFGDLINPFDTYKTPSKNIYTDNRKFFINEIKIEKKIESD 470
Db 173 -----NKLRELDRIEQDYNFRKKTVE-----EKARVLELKK--DREKLEDE 212
QY 471 KKSIEDRSKSLNDITKEYEKLLNELIYDSKFNNDILTNFEKMMGKRYKYVEKLTHTHN-T 529
Db 213 IKNEIKRIKIDKQDFDEYK-----KRNOYLKLTTLTKI-----KEGELNELRS 257
QY 530 FASYENSKHNLEKLTALKVMEYSLRNIVVEKELKYYK-NLISKIENEIETLVENIKKD 588
Db 258 IEELRKOTENMDQLEKEINELE--NLRNI-----KLKFEKYEVLAKSHTSEMANVINLEKE 311
QY 589 EQLFEKKI-TKDENKPDKEITILEVSDIVKVOQVLLMKNKIDELK-KTQLILK----- 639
Db 312 IEE-YEKAIRRKELEP--KYLKYE-----LERKLEELQPKYQVLLKLSLDLS 358
QY 640 --NV-----ELKHNHVPNSYKQENKQEPYLYLVKKEIDKLVKPMKVESLINEE 688
Db 359 KNLKERLEKDASELSNDIDKVNLSLEQVEE-----TRKKOLIN-LRAQLAKVESLISEK 411
QY 689 KKNIKTEGQSDNSPSTEGEITGOATKPGQAGSALSGDSVQAOAQKQAPQPPVPV 748
Db 412 NEII-----NNISQVEGE-----TCP--VCGRPLD-----EEHKOK-----II 442
QY 749 PEAKAQVTPPPAPVNNKTNVSK-----LDYLEKLYEFNTSYICHKYILVSHSTMNEKIL 804
Db 443 KEAKSYILQLELNKNELEELKKTITNELNKTREYRRLS-----NKKASYDNVM 491
QY 805 KOYKITEEESKLSLSCDPLDLFFNIQNNIPVMSWFDLSNLSQLFMFIEYKEMVCNLY 864
Db 492 RQLAKLNEETIENLHS--EIESLKNIDEI-----KKINEVEKELKL-YYEEMRLSKY 541
QY 865 KLKDNDKIKNLEAKKAVSTSVKTLSSSSMOPLSLTPQDKPEVSANDDTSHTNL--N 920
Db 542 TKEELDKRVRKLDKMKKKEIE-----KEMRGL-----ESELGLDKRALESKILDLEN 591
QY 921 NSLKFNILSLGKNKIYQELIG-----QKSSSENFYEK--ILKSDTFFYSEFTNFYKS 973
Db 592 KRVLDE-----MKKKKGILEDYIRQVKLLOEVKNLREEVNIQFDENRYNE-----LKT 642
QY 974 KADDIN-SLNDESKRK-KLEEDINKLKTLOLSFDLYNKYKLERLFDKKTIV----- 1025
Db 643 SLDAYNLSLKEKENRKSRIEGELESLEKIDIEISNRIANYELQLK--DREKIINAKNL 699
QY 1026 -----GKYKMOIKKLTLLKEQESKLSLNNPKHVQLNFVSFFNKKKEAEIAETNTL 1078
Db 700 EKIRSGALGERKLOQSYIITMTKQIENNLNDI-----ISKFDUSI 738
QY 1079 ENT--KILLKHYG-----LVKYNGESSPLKTL-----EESIQTED 1114

Db 1067 ERVKELEENVRKKAETSSLSKSLDEONLSVQLQKIKELQARIEELEEELEARNAR 1126
QY 593 --REK---KITDENKDPDEKILEVDIVKVQVQKVLNKNKIDELKKTQILKNVLEKHN 647
Db 1127 AKVEQRAELNDELGERLDEAGGATSAQIE--LNKKREAELLKIRDLLEASIQHEA 1184
QY 648 HVPNSYK--QENKQBPYILVLKKEIDKLKVPMPKVESLINEEK-----690
Db 1185 QTSALRKKHQDAANE-----MADQVDQLQ----KVSKLEKDKDLKREMDLLESQWTH 1234
QY 691 NIKTEGQSDNSPSPTEGITG--QATTKGQQAAGSALGDSVQAAQ-----EQKQAQPP 743
Db 1235 NMKNKGCEKVMKQFESQMSDLNARLEDSORSINELQSQSKRLQAENSDLTRQLEDAHR 1294
QY 744 VVPVPE---AKAQVPTTPAPVNNKNTENVKLDYLEKLYEFLNTSYICHYILVSHSTMN 800
Db 1295 VSVLSKERSQLSQLEDAARRSLEETRASKL-----QNEVRNMHADM- 1337
QY 801 EKILQYKTKREESKLSKSCDPLDILLFNQNNIPVWYSMFDS-----LNNLSQLPFM 852
Db 1338 DAIREQ--LEEQESK---SDVQRLSKANNEIQWRSKFSEGANRTTEEDQKRKLIG 1392
QY 853 EYKEMVCNLYKLDN--DKIKNLE-----EAKKVSTSVKTL-----SSSS 893
Db 1393 KLSEAEQTEAANAKCSALEKAKSLRQLEDMSTEVDANASVNMOKKQRAFDKTTAE 1452
QY 894 MOPSLTPODKPEVANDTSHSTNL---NNSLKLFIENLSLG-----KNKNYIYBELI--- 943
Db 1453 WQAKVNSLQSELENSQKESRGVSAELYRIKASIEYQD--SIGALRRKNKNLADEIHDIT 1510
QY 944 -----GOKSSE-----NYYE 953
Db 1511 DQSEGGSTHELDKARRLEMEKEELOALAEAEAGALEQEEBAKVMRAQLEIATVRNEID 1570
QY 954 KILKSDTFYNESFTNFVKSRADDINSLNDESK-----RKKLEEDINKLTKTLQLSF 1005
Db 1571 KRIQKEEDEFNTRRNHQRALESQASLEAEAKGAKADAMRTIKKLEQDINE-----LEVAL 1626
QY 1006 DLYNKYKLEKLEKLFKTKTVGKYKMQIKKL-----TLIKE 1040
Db 1627 DASNRGAEME-----KTVKRYOQIEMQTSIEEQORDEARESYNMAERRCTLMSG 1680
QY 1041 QLESKLSLNPKNPKVLQNFVSFFNKKEAEIAETENTLNTKILKHKYGLVKVYNGBSS 1100
Db 1681 EYEEELRAALEQAERA-----RRASDNLADANDRYNELTSQVSSVQGGQKRKLEGDIN 1732
QY 1101 PLKT-----LSEPSIQTEDNYASLENFVKLSKLECKLD- 1134
Db 1733 AMQTDLMHGLKGADECKKAMADAARLADLRAEQDHSNQVE--KVRKNLESQVKEF 1790
QY 1135 -----NLNLEKKLSVLSGLHHLIAEL-----KEVIKNKN-----YT 1167
Db 1791 QIRLDEAEASSIKGGCKMIQKLESVHELEALDNEQRHAETQKNMRKADRLKELAFQ 1850
QY 1168 GNSPSENN-----TDVNNALESYKFKLPEDGTAVT-----VSSSG 1203
Db 1851 ADEDRKNOERLQELIDKLNAKIKTFKQVEEAETIAINLAKYRKAQHELEAEERADTA 1910
QY 1204 SDTLQOSOPKPKPASTHVCAESNTFTS 1230
Db 1911 DSTLQKFRKSRSSVSQVRSSVSAS 1937

RESULT 58
MYSS_CHICK
ID MYSS_CHICK STANDARD; PRT; 1938 AA.
AC P13538; O13228;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, adult.
OS Gallus gallus (Chicken).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn; TISSUE=Pectoralis muscle;
RA Chao T.H., Bandman E., Moore L.;
RT "Cloning, nucleotide sequence and characterization of a full-length
RT cDNA encoding the myosin heavy chain from adult chicken pectoralis
RT major muscle.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 1-205.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041767; PubMed=1939027;
RA Hayashida M., Maita T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: I.
RT Sequence of the amino-terminal 23 kDa fragment.";
RL J. Biochem. 110:54-59(1991).
RN [3]
RP SEQUENCE OF 206-636.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041768; PubMed=1939028;
RA Komine Y., Maita T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: II.
RT Sequence of the 50 kDa fragment of subfragment-1.";
RL J. Biochem. 110:60-67(1991).
RN [4]
RP SEQUENCE OF 637-837.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041769; PubMed=1939029;
RA Maita T., Miyaniishi T., Matsuzono K., Tanioka Y., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: III.
RT Sequence of the 22 kDa fragment and the alignment of the 23 kDa, 50
RT kDa, and 22 kDa fragments.";
RL J. Biochem. 110:68-74(1991).
RN [5]
RP SEQUENCE OF 838-1938.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041770; PubMed=1939030;
RA Maita T., Yajima E., Negata S., Miyaniishi T., Nakayama S., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: IV.
RT Sequence of the rod, and the complete 1,938-residue sequence of the
RT heavy chain.";
RL J. Biochem. 110:75-87(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 1-808.
RX MEDLINE=87092420; PubMed=3467365;
RA Maita T., Hayashida M., Tanioka Y., Komine Y., Matsuda G.;
RT "The primary structure of the myosin head.";
RL proc. Natl. Acad. Sci. U.S.A. 84:416-420(1987).
RN [7]
RP SEQUENCE OF 842-1270.
RX MEDLINE=90121764; PubMed=2610940;
RA Watanabe B.;
RT "Complete amino-acid sequence of subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:1027-1034(1989).
RN [8]
RP SEQUENCE OF 852-1108.
RX MEDLINE=89374803; PubMed=2775482;
RA Watanabe B.;
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:549-558(1989).
RN [9]
RP SEQUENCE OF 1145-1270.
RX MEDLINE=89228549; PubMed=2713098;
RA Watanabe B.;
RT "Amino-acid sequence of the hinge region in chicken myosin
RT subfragment-2.";
RL Biol. Chem. Hoppe-Seyler 370:55-61(1989).
RN [10]

RP SEQUENCE OF 1957-1938 FROM N.A.
RX MEDLINE=87217964; PubMed=3034534;
RA Motilary D.M., Barringer K.J., Dodgson J.B., Richter H.E.,
RA Young R.B.;
RT "Genomic clones encoding chicken myosin heavy-chain genes.";
RL DNA 6:91-99(1987).
RN [11]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-843.
RX MEDLINE=93303624; PubMed=8316857;
RA Rayment I., Rypniewski W.R., Schmidt-Base K., Smith R.,
RA Tomchick D.R., Benning M.M., Winkelmann D.A., Wesenberg G.,
RA Holden H.M.;
RT "Three-dimensional structure of myosin subfragment-1: a molecular
RT motor.";
RL Science 261:50-58(1993).
CC -!- FUNCTION: MUSCLE CONTRACTION. MYOSIN IS A PROTEIN THAT BINDS TO
CC F-ACTIN AND HAS ATPASE ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U87231; AAB47555.1; -
DR EMBL; M16557; AAA48970.1; -
DR PIR; PX0050; PX0051.
DR PIR; A26821; A26821.
DR PIR; S02082; S02082.
DR PIR; S04501; S04501.
DR PIR; S05515; S05515.
DR PDB; 2MIS; 1I-JAN-97.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Phosphorylation; Acetylation;
KW Calmodulin-binding; Multigene family; 3D-structure.
FT INIT_MET 0 0
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT DOMAIN 783 812 IQ.
FT DOMAIN 838 840 HINGE.
FT DOMAIN 841 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 657 679 ACTIN-BINDING.
FT DOMAIN 759 773 ACTIN-BINDING.
FT MOD_RES 1 1 ACETYLATION.

FT	MOD_RES	35	35	METHYLATION (MONO-).
FT	MOD_RES	130	130	METHYLATION (TRI-).
FT	MOD_RES	551	551	METHYLATION (TRI-).
FT	MOD_RES	755	755	METHYLATION (MONO-).
FT	MOD_RES	697	697	ALKYLATION (SH-1).
FT	MOD_RES	707	707	ALKYLATION (SH-2).
FT	CONFLICT	907	907	C -> Q (IN REF. 7 AND 8).
FT	CONFLICT	980	980	L -> F (IN REF. 1).
FT	CONFLICT	1343	1343	E -> D (IN REF. 5).
FT	CONFLICT	1545	1545	S -> A (IN REF. 5).
FT	CONFLICT	1796	1797	HV -> QL (IN REF. 5).
FT	CONFLICT	1830	1830	S -> A (IN REF. 5).
FT	CONFLICT	1863	1863	I -> V (IN REF. 10).
FT	CONFLICT	1929	1931	IHG -> FH (IN REF. 10).
SQ	SEQUENCE	1938 AA; 223013 MW;	EDD01CEA2681E10F CRC64;	

Query Match 3.9%; Score 331; DB 1; Length 1938;
Best Local Similarity 19.1%; Pred. No. 0.00053;
Matches 241; Conservative 233; Mismatches 434; Indels 354; Gaps 56;

QY	237	KMEDYIK--KNKKTINELIEESKTTIDKNKNATKEEKKK--LYQAQYDLSIYNKQL	292
DB	837	KIKPLKSAESEKEMANKKEFEKTKELAKSEAKKELEKVMVLLQEKNDLQ---	QV 893
QY	293	EEAHNLISVLEKRIDTLKNE-----NIKELDKINEIKNPPSPANGTNPNTLLDNKKI	347
DB	894	QAEADSLADAERCQLIKTKLEAKIKVEERADEEE--INA-----ELTKKKRL	945
QY	348	EEHEKEIKEIAKTIFNIDSLFTDPLEYLLREKNKNIDISAKVE--TKESTEPNEYPNG	406
DB	946	ED---ECSELKKDI-----DLELPL-----AKVEKEKHATE-----	974
QY	407	VTPYLSYDINNALNELNSFGDLPDYTKPSKNIYTDNERKKFINIKIEKKIEKK	466
DB	975	-----NKVKNLTEEMAVLDETIACL--TKE-----KKALQEAHQ--TLDDLQ	1013
QY	467	ISDCKSYEDRSKSLNDITKEYEKLNEYDSKFNNDLTNFKEMGKRYKYVE---K	523
DB	1014	VEEDK-----VNTLTAKTKLEQQVDD--LEGSEQEKKLMDLERAKRKLEGGDK	1062
QY	524	LTHHTFASYENSKINLEKLTALKY-----MEDYSLRNIVVEKLYKKNLSKIE	575
DB	1063	LA-HDSIMDLENDKQOLDEKLAKKDFEISQISQKTEDEQALGMLOKKIKELQARIE	1121
QY	576	NEIETLVENIKKDEQLEFKITKDEKNKPEKILEYSDIVKQVQKVLMMNK--IDELKK	633
DB	1122	EEIE--AERTSRAKAEKHADLSRELEEFISERLEEAGGATAAQIE---MNKKREAFQK	1175
QY	634	TQILKNVELKH-----NIHVPNSYKQENKQEPYIYLVLKKE	670
DB	1176	MRRDLEATLQHEATAAALRKKHADSTABELGEQIDNLRVKQKLEKESK-----	1229
QY	671	ID-----KLKVPKPVESLINEEKKNIKTQESDSEPSSTEGEITQATKPGQ	720
DB	1230	IDDLASNMESVSKAKANLEKMCRTLEDQLSEIKTEEQNRMINDLNTORARLQTGET	1289
QY	721	AGSALGDSVQAQOEQKQAPVPVPVPEAKAQVTPPVPVNNKNTENYSKLD--YLEKLY	779
DB	1290	SQAEEKDALISQSRGKQ-----GFTQQIEELKRLHLEE--	1323
QY	780	EFLNTSYCHKYLYSHSTPMNEKILKYKITKEESKLSKSDPLDLLFNQNNIPVWSM	839
DB	1324	EIKAKNALAHALQSAHDC--ELLREQVE--EEQEA-----	1356
QY	840	FDLSLNNLSOLFWEI-----YEKEMVCNLYKLD--NDKIKNLEBAKKVSVKTLSS	892
DB	1357	-GELQALSKANSEVAQWTKYETDAIQRTTELEAKKLAQRLQDAEEHVAANA-KCA	1414
QY	893	SMQPSLTPODKPE---VSANDTSHSTNLNNSLKFENISLGKNKNYQEL-----	IG 944
DB	1415	SLEKTKQRLQNEVEDLWVDVRSNACALDKKQNFDKILAEWKOK--YEETQTELEAS	1472
QY	945	QKSS-----ENFYEKILKSDSTFYNESFTNFVKSADDINSLND-----	983

Db 1473 OKERSLSSTELFKMKNAYEESLDHLETKREN-----KNLQOEADLUTEQIAEGGKAVHE 1527
Qy 984 -ESRKKLEEDINKL-----KTLQSLDLYNKYKXLER-LFDKKTG 1026
Db 1528 LEKVKHVEQKSELQASLEAEASLEHBEKILRLQLEL-NQIKSEIDRKIAEKDEID 1586
Qy 1027 KYRM-QIKKLTLLKEQLESKLNPNKPHVLQNFVSFFNKKKEAEIAETENTLENTKILL 1085
Db 1587 QLRNHLRIVESMOSTLDAEIRS-----RNEALRLKKMEGDLNEMELQLSHANRMA 1638
Qy 1086 KHYGLVKYNGESSPLKTTSEESIQEDNVASLENKVLKLEGLKDNINLEKKLSY 1145
Db 1639 AEAQNLRLNTQG-----TLKDTQIHLDDALRTQED-----LKEQVAVMYERRANL 1682
Qy 1146 LSSGLHLIAELKEVINKNYTGNPSNNTDV-----NNALSYKKFLPEGTDVATVV 1199
Db 1683 LOAEVEELRGALQETRSKVAEQELDATERVOLLHTQNTSLNTKKLE--TDIVQIQ 1740
Qy 1200 SEGSSTLEQES-----QPKKPASTHWGAESNTITTSQ-----NVDDVEDDV----- 1240
Db 1741 SEM-EDIQEARNAEAKKAITDAAMMAELKKEQDTSAHLEKMKNMDOQTVADLHVL 1799
Qy 1241 -----IIVPIGESEEDYDGLGOVVTGEAVTSPVIDNLSKIENE 1280
Db 1800 DEABQALKGKKLOKLEARVRELEGEVDEQKRSAAEAVKGVKRYERRVKELTYQCEED 1859
Qy 1281 YE-VLYLKPLAG-----VYRSLLKQLENNVMTFNVNK-----DILNSRF 1319
Db 1860 RKNILRLQDLVKLOMKVYSKQAEAEELSNNLSKFRKIOHELEEAERADIAESQV 1919
Qy 1320 NK 1321
Db 1920 NK 1921

RESULT 59

TANA_XENLA STANDARD; PRT; 1744 AA.
AC Q01550;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tanabin.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUP-Tadpole head;
RX MEDLINE=92398961; PubMed=1524825;
RA Hemmati-Briantlou A., Mann R.W., Harland R.M.;
RT "A protein expressed in the growth cones of embryonic vertebrate
neurons defines a new class of intermediate filament protein.";
RL Neuron 9:417-428(1992).
CC -1- TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.
CC -1- DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURULA AND PERSIST
CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL
CC CORD.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; M99387; AAA49966.1;
DR PIR; JH0720; JH0720.

DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone.
FT DOMAIN 1 12 HEAD.
FT DOMAIN 13 314 ROD.
FT DOMAIN 315 1744 TAIL.
FT DOMAIN 8 48 COIL 1A.
FT DOMAIN 49 60 COIL 1B.
FT DOMAIN 61 156 COIL 1C.
FT DOMAIN 157 179 LINKER 12.
FT DOMAIN 180 193 COIL 2A.
FT DOMAIN 194 199 LINKER 2.
FT DOMAIN 200 314 COIL 2B.
SQ SEQUENCE 1744 AA; 199561 MW; 6502EAC9FE6C4E93 CRC64;
Query Match 3.9%; Score 330.5; DB 1; Length 1744;
Best Local Similarity 18.0%; Pred. No. 0.00049;
Matches 305; Conservative 311; Mismatches 574; Indels 509; Gaps 79;
Qy 132 VRNYLLTIKELKYPQLDNLTHMLTLCDNHGHKYLIDGYEEINELLYKLNFPDILRAK 191
Db 1 MEGYLASV-----SLGEESTQMWSLNRKLEAYLSRVKALEEENELLRK---EIHSLRSS 51
Qy 192 LNDVCANDYCOIPFNLRIRANELDLVKKLVFGYRKPL--DNIKDNVGMEDYIKKNKTI 249
Db 52 KSERC-----WKKKHHEEMMKLRDALDDGHEMVQAEWVRDSIYEIEFVK--QRCL 101
Qy 250 ENINELIEESKTTIDKNKNATKEEKKKLYOAYDLSYNNKQLEEAHNLISVLEKRI--DT 308
Db 102 EE-KOAREDAKKELESKLEETRAQIW-----LKER---LQGLEAELEDI 145
Qy 309 LKKNENIKELLDKINEIKNPPANSNGTPTLLDNKNKIEEHEKEIKETIKR-FNIDS 367
Db 146 LRDHEEKALM-----EEIASFSQRLENFRVAP 174
Qy 368 LFTDPLEYLYREKKNIDISAKVETKESTEPNEYPNGVTYPLSYNDINNALNELNSFG 427
Db 175 VAFKPVVDYAR-----KLSEIWQGAVEEYKSEVS-----LEAGLSE----- 213
Qy 428 DLINPFDTREPSKNITYDNERRKFFINEIEK-----IKIEKKKIES-DKKSYEDRSKSLN 482
Db 214 -----SKENLRKVLLENKONRLILQSLDKELVSLKMRKEALEDLSSKQWQEQEKEE 265
Qy 483 DITREKLLNIYDYSKFNINIDLTNPEKMMGKRYSYVEKLTHTHNTFASYENSKHNLEK 542
Db 266 KLOKRAEALEQKODLRQIAEVLEDROQLMHLKMSLEVATYRSLEA-----ES 317
Qy 543 LTRALKYMEDYSLRNIVVEKELYYKNLISKIENEIETLVENIKKDEQLEKFKITKDN 602
Db 318 TRIYTDYRGSYTFNDSMLE-----HNNVRRR---QSEDTKRTVSKDHRQSYKKQIGDN 369
Qy 603 ---KPD-----EKILEVSDIVKVOVKVLLMKNKIDELKKTQLILK 639
Db 370 ELQRPPLNFTSVKSSAVPVRTSPVTKEFQKVSQGLKYTKAPQKVEQVSTVTKS 429
Qy 640 NVELK-----HNIHVPSYKQENK-QEPPYL-IVLK 668
Db 430 NLEHTTSGDAFRAQVETRTDQVKKDALGLNDLNKNTGFKEEKDIOQPGMDHVS 489
Qy 669 KEIDKLVFMPKVESL-----INEEKNIKT---EGQSDNSEPSTEGETQATTKPGQ 719
Db 490 KSVSSTEHEKVPIDPLESALKSLEDLSVSVSTFNAGSSNLEAIK--DVLGPECLENL 547
Qy 720 QAGSALEDGVOAQAQEQKQAPVPVPEAKAQVTPPAPVNNKNTENSKLDYLEKLY 779
Db 548 QNEIAFEKESPGTNA-----AADPIEEVISES-----VSQTVHFHFKQE-LSNLL 591
Qy 780 EFLNTSYICHRYILVSHSTMNEKILKQYKITEESKLSSC--DPLDLLLFNIONNIPMY 837
Db 592 EIENT-----HE---NH-----VODTAQAFNSCEQDGHDRASTLENNEPDVQ 630
Qy 838 SMFDSLNNLSOLFMEIYEKEMVCNLYKDKNDKIKNLLEEAKKYSTSVKTLSSSSMQPL 897

Db	631	QYTRTLESN-----EIKESKIP-----SON-----TEAEIISKSRKVFLENEYIPV	672
Qy	898	SLTPQDKPEVSANDDTSHSTNLNNSLKLFEN-----ILSLGKN--KNIYQELIGQKSS	948
Db	673	S--KDDLTFEFTSHLENDSESSQPSDKLFENKSTEDQLITNLKSNTOENIFQS--NOEHL	728
Qy	949	ENF-YEKILKDSOTFYNESFTNFVYSK-----ADDINSLNDES-----	995
Db	729	ENLEFDSVDPDTVKFYPOENNLNLEENNVYGDGELVOMATDENTINQSSQLLSDHSHH	788
Qy	986	-----KRKKLEEDINKKLTQLSFDLKNKYKLG+-LERLPDKKKTGVKYMQI	1032
Db	789	EETKTSESTAVENRMSEHAZVDSKSEIPIVSEINVSVEIHIHESDVEEDT-----	841
Qy	1033	KKLTLLKEQLESKLNSLNNPKHVLQNFVFNK-----KKEAETAE--	1073
Db	842	-KQAFEDERVGEQINONQOESTVLDGSGVYSOEENSQLEDEVSISQIEKDFEINQECC	900
Qy	1074	-----TENTLENTKILLKHKGKLVKYNGESSPLKTLSEESI	1110
Db	901	LKSDQIREAPDTEEDVHQVDFPMQOSPEFREVGOINNIIQOEVDY-----LQNYDEDSF	953
Qy	1111	QTDNTYASLENFKVLKSEGLKMDLNEKKKLSYLSGLHLLAEKELKEVINKNNTGNS	1170
Db	954	QNDEPQOELESCDL--QBQKIK-----LEEN-----QISENEGQNFGGND	993
Qy	1171	PSNNTDYNNALESYKKFLPEGTDTAVTVSE-----SGSOTLEQ	1209
Db	994	IEE-----FSQOQYDTEICQETIGNOVSAQLLCESDINQDKLSMEQEEQ	1039
Qy	1210	SQPKPASTHVGAESNTITSON-----VDDEVDVVIIVPIGESEEDY-----DDLGOVVT	1261
Db	1040	NNPETEDNIGLQESQDENTRNECTKFSQBECDVVFVFPEDMSDKSEYSGQOEDLDKQVT	1099
Qy	1262	GEAVTPSDNTILSKTENEYVLYLKLPLAGVYRSLKLOLENNVTFNNVVKDITLNSRFNK	1321
Db	1100	DFSINEQANDLL--EKEEVILH--HADDORS-----VNDIEITIDEKLE	1140
Qy	1322	RENFKWVLSDLPIPKYKDLTSSNVVVKDYKFLNKEKRDKFLSSVNYIKDSIDTIDFAND	1381
Db	1141	R-----IIDNELATV-DVNES-----LAANKREQVDLFTDEY-----AVDONVGMQDD	1181
Qy	1382	VLGYYKILSEKYSKSDLSITKKY-INDKOGENEKYLPLFNNT-----ETLYKTV--NDKI	1432
Db	1182	DSQOYTQKEDLFDVGNNTIETKTEIQTOSLLNQETCERVDNVEDISEGAKNESVEMDVV	1241
Qy	1433	DLEVIHLEAKVLNVTYKSNVEVKIKELNYLTKTIQD-----KLADEFKN--NNFVGJAD-	1484
Db	1242	DLVP--EAKV--TGDEQISPLQDEKLN-LETMEDTKDNDGQCLCEKEKETEVIEVTD	1294
Qy	1485	--LSTDYNHN--NLLTKFLSTGWVP-ENLAKTVLSN-----LLDGNLOGLMNLISQHC	1532
Db	1295	PQFATDLSDHAGRELTVDQNSANLQFCENPTKTLIAHHIETVADSDLES---TBEQ-	1349
Qy	1533	VKKQCPQNSGCCFRHLDEREBCCLNLYQOEGDKCVENPNPTFCNENNGGCCDADAKCTEEDS	1592
Db	1350	-----VQETER-----IPFPEDSK-MENENSESESVDSQELSLSHSHKSEE	1390
Qy	1593	GSNGKKITCEKTPQSYPL	1611
Db	1391	FEISKDYQLEQTLPDVPTPL	1409

Search completed: March 31, 2003, 07:21:27
Job time : 124 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 07:20:46 ; Search time 38 Seconds
(without alignments)
4146.429 Million cell updates/sec

Title: US-09-269-874A-3

Perfect score: 8424

Sequence: 1 MKLIFLCSFLFFIINTQCV.....SNFLGISFLILMLILYSFI 1639

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8424	100.0	1639	2 S05603	major merozoite su
2	8191.5	97.2	1640	2 A24594	probable major sur
3	7965	94.6	1631	1 SAZOK1	major merozoite su
4	5136.5	61.0	1726	1 SAZQGM	major merozoite su
5	5121.5	60.8	1726	2 A45948	major merozoite su
6	5065	60.1	1701	2 A54498	major merozoite su
7	5053	60.0	1701	2 A26868	major merozoite su
8	3093.5	36.7	1751	2 A45604	major blood-stage
9	2999.5	35.6	1086	2 S16752	major merozoite su
10	2359	35.1	1726	2 A39401	merozoite surface
11	2829.5	33.6	1060	2 S06286	major merozoite su
12	2548.5	30.3	1772	2 A45532	major merozoite su
13	2509	29.8	1785	2 A45546	major merozoite su
14	2146	25.5	651	2 S47282	merozoite surface
15	2101	24.9	400	2 A45545	major merozoite su
16	1187	14.1	233	2 A25814	glycoprotein 185 -
17	1135	13.5	680	2 A28121	major merozoite su
18	1096.5	13.0	281	2 B54498	major merozoite su
19	1021.5	12.1	636	2 A45949	merozoite surface
20	815	9.7	168	2 A25522	major merozoite su
21	710.5	8.4	146	2 S03290	glycoprotein, 190K
22	659.5	7.8	138	2 P00125	major merozoite su
23	605.5	7.2	1939	2 T18372	repeat organellar
24	605	7.2	2401	2 T28676	rhostry protein -
25	568.5	6.7	2269	2 T28677	rhostry protein -
26	559.5	6.6	1979	2 C71622	hypothetical prote
27	529	6.3	115	2 B25120	major merozoite su
28	521.5	6.2	3394	2 T18501	hypothetical prote
29	507	6.0	2166	2 G70163	hypothetical prote

hypothetical colle
merozoite surface
transport protein
ORF MSV156 hypothe
reticulocyte-bindin
hypothetical prote
myosin-like protei
merozoite surface
merozoite surface
reticulocyte-bindin
serine/threonine-s
hypothetical prote
centromere protein
hypothetical prote
hypothetical prote
interaptin - slime
major merozoite su
surface-located me
conserved hypothet
Imp1 protein - Myc
hypothetical prote
myosin heavy chain
hypothetical prote
DNA polymerase - m
hypothetical prote
major merozoite su
phosphatase (acid
giantin - human
RAD2 endonuclease
hypothetical prote
major merozoite su
giantin - human
major merozoite su
hypothetical prote
hypothetical prote
major merozoite su
hypothetical prote
ovt1 protein - nem
myosin heavy chain
major merozoite su
hypothetical prote
major merozoite su
myosin-like coiled
probable membrane
probable nuclear p
RAD50 protein - ye
kinesin-related pr
probable membrane
hypothetical prote
nuclear migration
protein g377 - mal
hypothetical prote
chromosome segrega
hypothetical prote
major merozoite su
hypothetical prote
hypothetical prote
surface-located me
chromosome assembl
major merozoite su
DNA-directed RNA p
hypothetical prote
mature-parasite-in
hypothetical prote

103 365.5 4.3 2380 2 E71604
104 365 4.3 1690 2 T13030
105 364 4.3 1447 2 F82909
106 364 4.3 1642 2 T08880
107 362.5 4.3 1712 2 C71618
108 361.5 4.3 2244 2 F90563
109 361 4.3 3660 1 S02041
110 360.5 4.3 1714 2 F16009
111 360.5 4.3 2346 2 T13829
112 360.5 4.3 2708 2 T09079
113 359.5 4.3 4981 2 T18489
114 359 4.3 1271 2 A45555
115 358.5 4.3 2094 2 S31124
116 357.5 4.2 1817 2 H70179
117 356.5 4.2 971 2 A70179
118 356.5 4.2 1294 2 T18473
119 356 4.2 2599 2 F90608
120 355 4.2 1173 2 T43527
121 354 4.2 1997 2 F17607
122 353.5 4.2 2285 2 T12796
123 353 4.2 1169 2 A64505
124 352.5 4.2 1163 2 G97236
125 351.5 4.2 1935 1 S06006
126 351 4.2 2253 2 T30336
127 351 4.2 5005 2 F82884
128 350.5 4.2 1177 2 B75150
129 350 4.2 1802 2 G71616
130 349.5 4.1 1935 1 A37102
131 349.5 4.1 1940 2 A29320
132 349 4.1 1939 2 I48175
133 348 4.1 1532 2 T18438
134 348 4.1 2657 2 T18497
135 347 4.1 1225 2 A49464
136 346 4.1 1306 2 T28313
137 346 4.1 1819 2 A71928
138 346 4.1 1927 2 A59236
139 346 4.1 2819 2 T05080
140 345.5 4.1 1183 2 F90559
141 345.5 4.1 1927 2 G64585
142 345.5 4.1 2649 2 A40937
143 345 4.1 1313 2 A48467
144 345 4.1 1620 2 S61535
145 345 4.1 1817 2 D71606
146 345 4.1 1939 1 A46762
147 344.5 4.1 978 2 F07387
148 344.5 4.1 1830 2 E82909
149 344.5 4.1 1934 2 I48153
150 344 4.1 2136 2 A05037
151 343.5 4.1 1338 2 T18416
152 343.5 4.1 1902 2 C97702
153 343 4.1 1236 2 T18459
154 343 4.1 1367 2 T18466
155 342 4.1 2052 2 C97038
156 341.5 4.1 1269 2 F84730
157 341.5 4.1 1938 2 I49464
158 341.5 4.1 1979 1 S03166
159 341.5 4.1 2437 2 T18482
160 339.5 4.0 1787 2 G97222
161 339 4.0 100 2 C44866
162 338.5 4.0 1957 2 A45627
163 338.5 4.0 2364 2 I40884
164 338 4.0 1169 2 H70178
165 338 4.0 1938 1 S06005
166 337.5 4.0 1051 2 T18351
167 337.5 4.0 1164 2 T24806
168 336.5 4.0 1427 2 S22695
169 335.5 4.0 1272 2 C90593
170 335 4.0 1955 2 T30934
171 334.5 4.0 2677 2 A38194
172 334 4.0 1392 2 A43336
173 333.5 4.0 2178 2 S55805
174 333 4.0 1538 2 T29095
175 332.5 3.9 1935 2 A59286

hypothetical prote
microtubule bindin
hypothetical prote
NMDA receptor-bind
hypothetical prote
hypothetical prote
dystrophin, muscle
Ser/Thr protein ki
tpr homolog - frui
probable chloroqui
hypothetical prote
glutamate rich pro
tpr protein - huma
probable secreted
exodeoxyribonuclea
hypothetical prote
ABC transporter pe
sp8 protein - fliss
DNA helicase II BR
probable transglyc
P115 homolog - Met
ATPase involved in
myosin beta heavy
nuclear/mitotic ap
hypothetical prote
chromosome segrega
hypothetical prote
myosin beta heavy
myosin heavy chain
myosin heavy chain
hypothetical prote
hypothetical prote
hypothetical prote
ORF MSV152 probabl
cag island protein
embryonic muscle m
probable chloroqui
conserved hypotet
cag pathogenicity
bullous pemphigoid
myosin heavy chain
nucleotide-binding
hypothetical prote
myosin alpha heavy
conserved hypotet
myosin heavy chain
hypothetical prote
cell surface antig
hypothetical prote
hypothetical prote
phage-related prot
probable myosin he
alpha cardiac myos
myosin heavy chain
hypothetical prote
hypothetical prote
major merozoite su
myosin heavy chain
cytotoxin L - Clos
exodeoxyribonuclea
myosin alpha heavy
impl protein - Myc
hypothetical prote
resin - human
hypothetical prote
myosin-like protei
desmoplakin I - hu
microtubule-vesicl
alpha-toxin - Clos
cardiac muscle fac
myosin heavy chain

176 332 3.9 1883 2 G82875
177 331.5 3.9 864 2 B90395
178 331.5 3.9 1938 1 A40997
179 331 3.9 3685 1 A27605
180 330.5 3.9 1744 2 JH0720

ALIGNMENTS

RESULT 1

S05603
Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
Submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from P1
A:Reference number: S04850; MUID:89345116; PMID:2668887
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 100.0%; Score 8424; DB 2; Length 1639;
Best Local Similarity 100.0%; Pred. No. 1.5e-253;
Matches 1639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTIIFLCSPFLTIINTQCVTHESYQLVKKLEALEDAVLGTSLFOKEKWLNEGTS 60
Db 1 MKTIIFLCSPFLTIINTQCVTHESYQLVKKLEALEDAVLGTSLFOKEKWLNEGTS 60
QY 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGNSRRTNPSDSS 120
Db 61 AVTTSTPGSKGVSASGGSGSVASGGSGSVASGGSGSVASGGSGNSRRTNPSDSS 120
QY 121 DAKSYADLKHRVRYLLTITKELYPQLFDLTNHLMLCDNIHGFYKYLIDGYEINELLYK 180
Db 121 DAKSYADLKHRVRYLLTITKELYPQLFDLTNHLMLCDNIHGFYKYLIDGYEINELLYK 180
QY 181 LNFYFDLLRAKLDVNCANDYCOQIPFNKIRANELDVLKLVFGYRPLDNIKNVGMED 240
Db 181 LNFYFDLLRAKLDVNCANDYCOQIPFNKIRANELDVLKLVFGYRPLDNIKNVGMED 240
QY 241 YIKKNKTTIENINELIEESKTTIDKNKNATKEBEKKLYQAOYDLSYNNKOLEEAHNLIS 300
Db 241 YIKKNKTTIENINELIEESKTTIDKNKNATKEBEKKLYQAOYDLSYNNKOLEEAHNLIS 300
QY 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANGTNPNTLLDNKKKEEHEKEIKETAKT 360
Db 301 VLEKRIDTLKKNENIKELLDKINEIKNPPANGTNPNTLLDNKKKEEHEKEIKETAKT 360
QY 361 IKENIDSLFDPLELYLREKNKNIDISAKVETKESTNEYPNGVTYPLSYNDINNAL 420
Db 361 IKENIDSLFDPLELYLREKNKNIDISAKVETKESTNEYPNGVTYPLSYNDINNAL 420
QY 421 NELNSFGDLINPFDYTKEPSKNITYDNERKKFNEIKKIKIEKKIESDKSYEDRSKS 480
Db 421 NELNSFGDLINPFDYTKEPSKNITYDNERKKFNEIKKIKIEKKIESDKSYEDRSKS 480

hypothetical prote
purine NTPase [imp
myosin heavy chain
dystrophin, muscle
tanabin - African

Qy	781	FLATSYICHXYILVSHSTWNEKILKOYKITKEEBSKLS	CDOPLDLLFNQNNIPVMTSMF	840
Db	772	FLATSYICHXYILVSHSTWNEKILKOYKITKEEBSKLS	CDOPLDLLFNQNNIPVMTSMF	831
Qy	841	DSLNNLSQLFMIEYKEWCMVNLYLKDNDK1KNLLEAKK	VYSTVKTLSSSSWOPLSLT	900
Db	832	DSLNNLSQLFMIEYKEWCMVNLYLKDNDK1KNLLEAKK	VYSTVKTLSSSSWOPLSLT	891
Qy	901	PODKPEVSANDDTSHSTNLNLSKLNFENILSLGKNKIYQ	ELIGOKSSSENEYEKILKDSO	960
Db	892	PODKPEVSANDDTSHSTNLNLSKLNFENILSLGKNKIYQ	ELIGOKSSSENEYEKILKDSO	951
Qy	961	TFYNESFTNPFVSKADDIINSLNDESKRKKLEEDINKLK	TKTQLSPDLNKKYKLERLFD	1020
Db	952	TFYNESFTNPFVSKADDIINSLNDESKRKKLEEDINKLK	TKTQLSPDLNKKYKLERLFD	1011
Qy	1021	KKTVGKYKMOIKKTLLEKQELSKLSNLNPNKPHVLQNF	SFFNKKKEABEATENTLEN	1080
Db	1012	KKTVGKYKMOIKKTLLEKQELSKLSNLNPNKPHVLQNF	SFFNKKKEABEATENTLEN	1071
Qy	1081	TKILLKHXYGLVKNYNGESSPLKTISEESIQTEDNYSLE	NFKVLKLEGLKDNLNALEK	1140
Db	1072	TKILLKHXYGLVKNYNGESSPLKTISEESIQTEDNYSLE	NFKVLKLEGLKDNLNALEK	1131
Qy	1141	KKLSYLSLGLHLLIAELKEVINKNYTGNSPENNTDVNN	ALESYKKFLPGTDVATVVS	1200
Db	1132	KKLSYLSRGLHLLIAELKEVINKNYTGNSPVNNTDVNN	ALESYKKFLPGTDVATVVS	1191
Qy	1201	EGSDTLEQSQPKPASTHWGAESENTTTSQNVQDVEDDV	VIIVPIFGSEEDYDGLGOVY	1260
Db	1192	EGSDTLEQSQPKPASTHWGAESENTTTSQNVQDVEDDV	VIIVPIFGSEEDYDGLGOVY	1251
Qy	1261	TGBAVTPSVIDNLTLSKIENEYEVLYLKLPLAGVYRS	LKKOLENNVMVTFNVYKDLNSRFN	1320
Db	1252	TGBAVTPSVIDNLTLSKIENEYEVLYLKLPLAGVYRS	LKKOLENNVMVTFNVYKDLNSRFN	1311
Qy	1321	KRENFKNVLESOLIPYKDLTSSNYVVKDOPYKFLNKE	KRDKFLSSVNYIKDSIDTDINFAN	1380
Db	1312	KRENFKNVLESOLIPYKDLTSSNYVVKDOPYKFLNKE	KRDKFLSSVNYIKDSIDTDINFAN	1371
Qy	1381	DVLGYTKILSEKYSKDLDSIKKYINDKOGENEKYLPP	FLNNIETLYKTVNDKIDLFTVHLE	1440
Db	1372	DVLGYTKILSEKYSKDLDSIKKYINDKOGENEKYLPP	FLNNIETLYKTVNDKIDLFTVHLE	1431
Qy	1441	AKVLNTYSEKSNVEVKIKELNYLTKTIDQKLADF	KKNNFVGIADLSTDYHNHLLTKFLS	1500
Db	1432	AKVLNTYSEKSNVEVKIKELNYLTKTIDQKLADF	KKNNFVGIADLSTDYHNHLLTKFLS	1491
Qy	1501	TGMVFENLAKTVLSNLLDGNL-OQGMINTISOHCYKQ	CPQNSGCFRHLDEREECKCLLNY	1559
Db	1492	TGMVFENLAKTVLSNLLDGNL-KARYVKKHFTPMRK	KTKMTIQQNSGCFRHLDEREECKCLLNY	1551
Qy	1560	KQBGDKCVENPNTCENNINGGDADAKCTEEDSGSNGK	KITCECTKPDSPYLPFGIFCSS	1619
Db	1552	KQESKCVENPNTCENNINGGDADAKCTEEDSGSNGK	KITCECTKPDSPYLPFGIFCSS	1611
Qy	1620	SNFLGTSFLLILMLILYSPF	1639	
Db	1612	SNFLGTSFLLILMLILYSPF	1631	

SECRET

RESULT 4
 SAQGM
 Major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
 N: Alternate names: 195K glycoprotein
 C: Species: Plasmodium falciparum
 C: Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
 C: Accession: A23386; S06361
 J: Weber, J.L.; Leininger, W.M.; Lyon, J.A.
 Nucleic Acids Res. 14, 3311-3323, 1986
 A: Title: Variation in the gene encoding a major merozoite surface antigen of the human m
 A: Reference number: A23386; MUID: 86205236; PMID: 3517809

Qy	678	MPKVESLINEKNIKTEGQSDNSPSETEGETQATTKPQQQAGSALLEGDSVOQAQOEQ	737
Db	710	IPKVKMLKKEQAVLSS-----ITQPLVAASSETTEGDGTHHTLSOSGETEVEFTE	760
Qy	738	KOAOQPPV-----PVPVPEAKAOPVPPAPVNNKKTENVSKLDYLEKLYEFLNTSVICH	789
Db	761	EETEEVVGHTTVTTITLPPKEVKVYVENSIEHKSNDNSQALTKTVYLAKKDEFLTKSVICH	820
Qy	790	KYILVSHSTMNEKILKOYKITKEESKILSCDPLDLLFENIQNNIPVMYSFMFDSLNNLSQ	849
Db	821	KYILVSNSSMDQKLEEVNLTPEENELKSCDPLDLLFENIQNNIPAMYSVLVDSMNNDLQH	880
Qy	850	LFMEIYEKWCVNLKYLKNDKIKNLLBEAKV-----STSVKTLSSSS-----	893
Db	881	LPFELYOKEMIIYHLKLBENHIKLLBEQKQITGTSTSSPGTNTVNTAQSAHNSQN	940
Qy	894	--MQLSLTPQOKPEVSAND--DTSHS-----TNLNNLSKLFEILSLGNKNKYQEL-I	943
Db	941	QOSNASSTWQNGVAVSSGPAVVEESHDPJTVLSISNDLKGIVLSLLNLGNKTKVPNPLTI	1000
Qy	944	GQKSSSENFYEKILKQSDTPYNSFNTFVKSRADDINSUNDESKRKKLEEDINKLTKTLQ	1003
Db	1001	STTEMEKFYENILKNNDTYFNDIDIKQFVKSNSKVITGLT-ETQKNALNDEIKKLDPQLQ	1059
Qy	1004	SPDLYNKYKLERFLDKKTKGVKMQIKKLTLLKQELSKLASLNNPHVLQNSVFF	1063
Db	1060	SPDLYNKYKLLDRLPNKKKEGQDKMQIKKLTLLKQELSKLASLNNPHVLQNSVFF	1119
Qy	1064	NKKKEAIEAETENTLNTKILKHKYKGLVKNYNGESSPKLTLSVESIQTEDNVANLEKPR	1179
Db	1124	VLSKLECKLDNLNLEKXKLSYLSGLLHLLTAELKEVTKNKNYTCNPSENNVDVNNALE	1183
Db	1180	VLSKIDGKLDNLHLGKKLSFLSGLLHLLTELKEVTKNKNYTCNPSENNKKVNEALK	1239
Qy	1184	SYKFKLEPGTDVATVYSE-----SGSDTLEQSQPKKPASTHVGAESNTITT	1229
Db	1240	SYENFLPE-AKYTTVTVPPQDPVTPLSVRVSCSGSKTEETOIPTSGSLLTQLQOVQ	1298
Qy	1230	QSNWDEVDDVIIVPIFGSESEEDYDDLGQVTVGEAVTPSVTDNLISKIENEYVLYLKPL	1289
Db	1299	LQNYDEEDSLVLPPIFGSEDNDEYLDQVVTGEAISVT-MDNILSGFENEYDVIYLLKPL	1357
Qy	1290	AGYVRSLLKOLENNVMTFNVNKDILNSRENKRNFKNVLDESILPYKDLTSSVYVKDP	1349
Db	1358	AGYVRSLLKQIEKNTFTFNLNDILNSRLKKRYFLDVLDESILMQPKHSSNEYIIEDS	1417
Qy	1350	YKFLNKEKRDKFLSSNYIKOSIDPTDINFANDVLGYKILSEKYKSLDLSIKKYI-----	1404
Db	1418	FKLLNSEQNTLLKSYKYIKESVENDIKFAQEGISYEKVLAKYKDDLESITKKVKEBE	1477
Qy	1405	-----NDKGENEKYLPFLNLTETLYKTVNDKIDLFVHLEAKVLNYYE	1449
Db	1478	KFPSSPTTPPGAKTDEQKESKEFLPTNIETDYNLWNKIDDYILINLAKINCQVE	1537
Qy	1450	KSNVEYKIKELNYLKTIOQKLADFKNKNFVGADLSPTDYNNHLLTKFLSTGMVFENLA	1509
Db	1538	KDEAHVKITKLSDLKAIDDKIDLPKNHNDFAIKKLINDDTFKDMLGKLLSTGLV-QNFP	1596
Qy	1510	KTVLSNLLDGNLQGMINTSIOHCYKQCPQNSGCFRHLDEREBCKLLNTYKQEGDKCVEN	1569
Db	1597	NTIISKIEGKFQDMLINTSIOHCYKQCPQNSGCFRHLDEREBCKLLNTYKQEGDKCVEN	1656
Qy	1570	PNPTCNENGGCDAKACTEEDSGSNKKITCECTKPDSPYPLFGIFCSSNFLGTSFLL	1629
Db	1657	PNPTCNENGGCDAKACTEEDSGSNKKITCECTKPDSPYPLFGIFCSSNFLGTSFLL	1716
Qy	1630	ILMLIYLSFT	1639
Db	1717	ILMLIYLSFT	1726

RESULT 5
A45948 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C:Accession: A45948
R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A>Title: Plasmodium falciparum: gene structure and hydropathy profile of the major me
A:Reference number: A45948; MUID:89005525; PMID:3049134
A:Accession: A45948
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726 <CHA>
A:Cross-references: GB:M37213
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match	60.8%	Score 5121.5;	DB 2;	Length 1726;
Best Local Similarity	60.0%;	Pred. No. 2.3e+151;		
Matches 1050;	Conservative 221;	Mismatches 344;	Indels 135;	Gaps 26;

Qy	1	MKIIFLCSPFFLIINQCVTHTSYQLVKKLGALEDAVLGTGSLFQKMKWLNE-----	55
Dd	1	MKIIFLCSPFFLIINQCVTHTSYQLVKKLGALEDAVLGTGLFHKKMWLNIEEITT	60
		: :	:
Qy	56	-----GTSQTAVTTSTPGSKGVASGGSGSVASGSVASGSVSAGS	105
Dd	61	KGASAQSGETSGTSGTSGTSAOISGTSQAOSGT--SGTSAQSGETSGTSGT	118
		: :	:
Qy	106	GN---SRRTN-----PSDNSSDASKYADLUKHVRNVLTITIKELKYQPDLDT	151
Dd	119	GTSPSSRSNTLPNSTSSGASPADAADSADAKYADLUKHVRNYLFTEIKELYPELDET	178
		: : :	:
Qy	152	NHMLTLCDNIHGPKYLIDGYEEINELLYKLNFYFDLLRAKLNDVCANDYOIENLKIRA	211
Dd	179	NHMULTCDNIHGFKYLIDGYEEINELLYKLNFYFDLLRAKLNDVCANDYOIENLKIRA	238
		: :	:
Qy	212	NELDVKLVFGYRKPLDNINKVNKGMEYDIKNKTITIENIELIESKTKTDKNATK	271
Dd	239	NELDVKLVFGYRKPLDNINKVNKGMEYDIKNKTITIANIELIEGSKTIQNKNADN	298
		: :	:
Qy	272	EEEKKKLYAQAYDSITYNKQIEAHNLISYLEKRIDLTKNNENIKELLDKINEIKPPPA	331
Dd	299	BEGKKLYAQAYDSLTYNQKEEAHNLSILEKRIDLTKNNENIKELLDKINEIKPPPA	358
		: :	:
Qy	332	NSGMTPTNLDDNKKIEEHKEIKEIAKTIKFNIDSFTDPLEYLYREKKNINDISA	391
Dd	359	NSGMTPTNLDDNKKIEEHKEIKEIAKTIKFNIDSFTDPLEYLYREKKNKVDTWK	418
		: :	:
Qy	392	VE--TKESTEPN-EYPNGVTYPLSYNDINNAL---NELNSFGDLINFDFYTKEPSENIYT	445
Dd	419	SODPTKSQVIPKPVPYPIINGIVPLPLTDIHNSLAADDKNSYGDMNP-DTKENEKIIT	477
		: : :	:
Qy	446	DN-ERKKFIENIEKIEKIIKKI----ESDKSYSEDRSSLNDITKEYEKLLNEIYDSKF	500
Dd	478	DNKERKIFIANNIKOIDLEEKKINHTEQQNKILLEDEYS-----KDYEELLEEFYMFP	533
		: : :	:
Qy	501	NNIDTLTFNMCKMGRYSVKEVELTHHTNTFSYENSXHNLEKLTALKTKMYEDSLRNIV	560
Dd	534	NNPDFDVVDKIPARSARTYNVEQRYYNNKSSSNNSVYNQVKKALSILEDYSLRGIS	593
		: :	:
Qy	561	EKELYKNILISKTENBIETFLVENIKKDEBOLFEKK---ITKDENVKDELIVSDIVKV	617
Dd	594	EKDNHYVTVTLTGLEDIKILLTBKISSSENKILEKNFKGLTHSANAS----LEVYDIVKL	649
		: :	:
Qy	618	VQVQVLLMNKIDELKKTQLLTKNVELKHNITHPVNSYKQENKOPEYVILVLKIEDLIKVF	677
Dd	650	VQVQVLLIKIEDLRTELFLKNAQLKDSLHVNIYKPNKPEPYVILVLUKREVCLKRF	709
		: :	:
Qy	678	MPVESLINEEKNIKETGSDNSEPSTEIGEITGCATTCPGOAGSALGEDSVQAOAEOQ	737
		: :	:

Db 710 IPKVDMKKKEQAVLSS-----ITQPLVAASETTEDGGHSTHTLSQSGETEVTET 760
Qy 738 KOAPV-----PVPVPEAKAQPPTPPAPVNNKNTENSVKLDYLEKLYEFLNTSYICH 789
Db 761 ETEETVGHVTTVTITLPPKEVKVYVNSLEHKSNDNSQALTTVTYLLKLUDEFLTYSYICH 820
Qy 790 KYILVSHSTMNKILKYKITKEESKLSKSCDPLDLLFNQIANNIPVYSMFDSLNSLSQ 849
Db 821 KYILVSNSSMDQLLEVYNTPEENELKSCDPLDLLFNQIANNIPAMYSLYDSMNDLQH 880
Qy 850 LFWEIYEKEMVNLKYLKNDKIKNLEBAKV-----STSVKTLSSS----- 893
Db 881 LFFELYQKEMYYLHKLEHNKILKEQKQITCTSTSSPCNTVTNPAQATHNSON 940
Qy 894 --MQPLSLTPQKPEVSAND---DPSHS---TNLNSLKLFPENILSLGKNKIYOEL-I 943
Db 941 QOSNASSTWONGVAVSSGPAVVEESHDPDLTVLSISNDLKGIVSLNLGNKTKVNPPLTI 1000
Qy 944 GOKSENFEYKILKSDTPTYNESFTNFVYSKADDINSLNDESKRKKLEEDINKLKTQL 1003
Db 1001 STTEMEKFENTLKNNDTVFNDIDQFVKSNSKVITGLT-ETQKNALNDEIKKLDTLQL 1059
Qy 1004 SFDLNYKYLKERLFDKKTGKYKQWKILTLLEQLESKLSLNNPKHYLQNFVFF 1063
Db 1060 SFDLNYKYLKLDRLFNKKELGQDMQIKLTLLEQLESKLSLNNPHNYLQNFVFF 1119
Qy 1064 NKKKEAETAEENTLTKILLKHVKGLVYNGESSPLKTLSESIQEDNYASLENPK 1123
Db 1120 NKKKEAETAEENTLTKILLKHVKGLVYNGESSPLKTLSESIQEDNYANLERER 1179
Qy 1124 VLSKEGKLUKNLEKKLSYLSGLHLIAELKEVINKNKYTGNSPENTDNNNALE 1183
Db 1180 VLSKIDGKLDNLHLGKKLSLSSGLHOLITELKEVINKNKYTGNSPENKNKYNEALK 1239
Qy 1184 SYKKFLPECTDVATVYSE-----SGSDTLQSQPKKPASTHVGAESNTIIT 1229
Db 1240 STENLPE-AKVTVVTPQPDVTPSPLSVRVSGSGSTKEETQIPTSGSLTELOQVVQ 1298
Qy 1230 SONVDEVDVIVIPFGESEDDYDQGVVTEAVTPSVIDNLSKIENEYEVLYLKLPL 1289
Db 1299 LQNYDEEDSLVPLPFGESDNDYLDQVVTGEAISVT-MDNILSGFENEYDVLYLKLPL 1357
Qy 1290 AGYVSLKQLENNVWTFNVVVKDIILNSFRKNRENFKNVLESDLIPYKDLTSSNVVVKDP 1349
Db 1358 AGYVSLKQKIEKNFTFNLNLDILNSRLKRRKYFLDLESQKFKHISSNEYIIDS 1417
Qy 1350 YKFLNKRKDRKFLSVNYIKSDSIDTDINFANDVLGYKITLSKYKSDLSIKKYI----- 1404
Db 1418 FKLLSEQNTLLSKYKIKESVENDIKPAQBGISYIEKVLAKYKODLESIRKVIKEKE 1477
Qy 1405 -----NDKOGENEKYLPLFNNTIETLYKTVNDKIDILFVHLEAKVLNITYE 1449
Db 1478 KFPSSPPTTTPSPVKTEQKESKFLPFLTNIETLYNLVKNDDYLLINKAKINDCNVE 1537
Qy 1450 KSNVEVKIKELNYLKTIOQKLADFKNNFVGIADLSTDYNNHLLTFLPSTGMFENIA 1509
Db 1538 KDEAHVYKTKLDLKAIDDKIDLFKNHNDFEAIKLLINDDTKMDLGLKLLTGLV-QNPP 1596
Qy 1510 KTVLSNLDGNLQGMNLISQHCQKQPCQNSGCFRHLDERECCLLNYKQEGDKCVEN 1569
Db 1597 NTIISKLIEGKFQDMLNISOHCQKQPCQNSGCFRHLDERECCLLNYKQEGDKCVEN 1656
Qy 1570 PNPTCENNGGDAADAKCTEEDSGNGKKTCTCTKPDSPYPLFDGIFCSSNPLGTSFLL 1629
Db 1657 PNPTCENNGGDAADAKCTEEDSGNGKKTCTCTKPDSPYPLFDGIFCSSNPLGTSFLL 1716
Qy 1630 ILMILYISFI 1639
Db 1717 ILMILYISFI 1726

RESULT 6
A54498

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmo
C:Reference number: A54498; MUID:88142999; PMID:2449612
A:Accession: A54498
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PET>
A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 60.1%; Score 5065; DB 2; Length 1701;
Best Local Similarity 60.1%; Pred. No. 1.3e-149;
Matches 1042; Conservative 227; Mismatches 338; Indels 126; Gaps 29;
Qy 1 MKIIIFLCSFLFIINTQCVTHESYOELVKLEALEDAVLTVGSYLFQKMKVNLNCTSGT 60
Db 1 MKIIIFLCSFLFIINTQCVTHESYOELVKLEALEDAVLTVGSYLFQKMKVNLNCTSGT 60
Qy 61 AVTTTPGSKGVSAGSGSGSVASGSGSVASGSGSVASGSGSGSNRRTNPDSNDS 120
Db 61 AVTTTPGSGSVT---SGGSVASVASVSG---SGGSVASGSGSNRRTNPDSNDS 114
Qy 121 DAKSVADLKHVRNVLITIKELYPQLFDLTNHLMTLCDNIHGFKYLIDGYEINELLKY 180
Db 115 NTKTYADLKHVRNVLITIKELYPQLFDLTNHLMTLSKNVDFKYLIDGYEINELLKY 174
Qy 181 LNFYDLLRAKLVNDVANDYCOIPNKLIRANELDVLKLVGYRKLPLNIDKNVGMED 240
Db 175 LNFYDLLRAKLVNDVANDYCOIPNKLIRANELDVLKLVGYRKLPLNIDKNVGMED 234
Qy 241 YTKNKKTTIENELIEESKKTIDKNKNATKEEKKLYQAOYDLSIYNKQLEEAHLNIS 300
Db 235 YTKNKKTTIANTIELIEGSKKTIDQKNADNEEGKKLYQAOYVNFYINQKQLEEAHLNIS 294
Qy 301 VLEKRITDLTKNENIKELDKINELK---NPPANSNGTNPNTLLD-KNKKTEEHEKEIK 355
Db 295 VLEKRITDLTKNENIKELDKINELK---TTGSKPAPLPENKKKEGHEEIK 351
Qy 356 EIAKTIKFNIDSLFDPLELYYLRKKNKIDISAKVE--TKESTEPN-EYPNGVYPLS 412
Db 352 EIAKTIKFNIDSLFDPLELYYLRKKNKVDVTPKSQDPTKSQVQPKVPYNGIYVPLP 411
Qy 413 YNDINNAL---NELNSFGDLINPFDYTKEPSNIYTDN-ERKKFNIKEIKIEKKKI- 467
Db 412 LTDIHNLSAADNDKNSYGDLMNP-DTKEKINEKIITDNKERKIFINNIKKQDLBEKNIN 470
Qy 468 ---ESDKSYEDRSLSNDITKEYEKLLNEIYDSFNNDNIDLTNFEKMMGRKYSYKVEL 524
Db 471 HFKQNKKKLLEDEYS---KKDYELLEKFEMFNNFNFDKVDVDFKISARTYTNVEKQ 526
Qy 525 THHNTFASYSKHNLEKLTAKLYMEDYSLRNIVVEKELKYKNLISKIEINIEITLVEN 584
Db 527 RYNNKFNSSNNVYNQKLLKALSLEYDSLGRKISEKDFNHYTLTKTGLEADIKKLTTEE 586
Qy 585 IKKDEEQLEFEK---ITKDNKPDEKILEVSDIVKVOQVLLMKNKIDELKTKQLILKNV 641
Db 587 IKSSNNKILEKNFKGLTHSANAS---LEVSDIVKLQVQVLLIKKIEDLRKLTFLKNA 642
Qy 642 ELKHNHVPNSYKQENKQEPYLYLVLLKEDKLKVPKPVKESLINEEKNKIKTEGSDNS 701
Db 643 QLKDSIHVPNIYKPNKPEPYLYLVLLKEDKLKPEIPKVKOMLKEQAVLSS----- 695
Qy 702 EPSTEGETTQATTKPGQOAGSALGSDSVQAOAQEQKA-----QPPVPVPEAKAQ 754
Db 696 --ITQPLVAASETTEDGGHSTHTL-SQSGETEVTETVGHVTTVTITLPPKEES 752
Qy 755 VPTPPAPVNNKTN-----VSKLDYLEKLYEFLNTSYICHYILVSHSTMNEKILQ 806


```
Db 753 APREVKVWENSIEHKSNDNSQALTKTVYKLLKDEFLTSTKSYCHKYLIVSNSSMDQKLEY 812
QY 807 YKLTKEESKLSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFMEIYEKEMVKNLYKL 866
Db 813 YNLTPEENELKSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFMEIYEKEMVKNLYKL 872
QY 867 KMDKIKNLEAKV-----STSVKTLSSSS-----MQLSLTLPQDKPEVS 908
Db 873 KEENHIKLLLEQKQITGSTSSPGNTVNTAQSATHNSNQOQSNASTNTQNGVAVS 932
QY 909 AND---DTSHS---TNLNNLSKLFENILSLGKKNKIYQEL-IGQSSSENFYEKILKDS 960
Db 933 SGPAVVEESHDPVLTSISNDLGIIVSLNLGNKTKVPNPLTISTTEMEKFYENILKND 992
QY 961 TFYNESFTNPKSKADDDINSNDSEKRLKEEDINKLKTLOLSPDLYNKYLKLERLD 1020
Db 993 TYFNDDIKOFVNSKVITGLT-ETOKNALNDEIKLKDTLOLSPDLYNKYLKLERLD 1051
QY 1021 KKTGVGKYMQIKKLLKLEQLESKLSLNNPKHVLQNFVFPNKKKEAEIAETENTLEN 1080
Db 1052 KKELGQDKMQIKKLLKLEQLESKLSLNNPHVNLQNFVFPNKKKEAEIAETENTLEN 1111
QY 1081 TKILLKHYGLVYNGESSPLKTLSEESTQTEDNYASUENFVLSKLEGLKDNLENLEK 1140
Db 1112 TKILLKHYGLVYNGESSPLKTLSEVSTQTEDNYANLEKPRALSCKIDGKLDNLEHLG 1171
QY 1141 KLSYLSLGGHLHLIAELKEVINKNNTGNSPSENNTDVNNALESYKFLPEGTDVATVVS 1200
Db 1172 KLSYLSLGGHLHLITELKEVINKNNTGNSPSENNTDVNNALESYKFLPEGTDVATVVS 1230
QY 1201 E-----SGSDTLEQSPKPKASTHVGAESNTITTSQNVDDVEDDVIIVPIF 1246
Db 1231 PPQPDVTPSLSVRVSNGSGSTKEEQIPTSGLLTELQVQVQLQNYDEDDSLVLP 1290
QY 1247 GESEEDYDLQGVVTEGATPSVIDNLSKIENYEVYIKPLAGVYRSLLKOLENNVT 1306
Db 1291 GESEEDNEDYDLQGVVTEGATPSVIDNLSKIENYEVYIKPLAGVYRSLLKOLENNVT 1349
QY 1307 FNVNVDILNSRKNRENKFNKLVESLDIPYKDLTSSNVYVQDPYKFLNKEKREKLSYNN 1366
Db 1350 FNVNVDILNSRKNRENKFNKLVESLDIPYKDLTSSNVYVQDPYKFLNKEKREKLSYNN 1409
QY 1367 YIKSDITDINFANDVGLYKILSEYKISDLSIKYI-----ND 1406
Db 1410 YIKESVENDIKFAQEGISYEYKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTD 1469
QY 1407 KQGENEKYLPFLNNIETLYKTVDNKDLDLFIHLEAKVLNITYEKSNNVEYKIKELNLYKTI 1466
Db 1470 EQKESKFLPFLNIETLYKTVDNKDLDLFIHLEAKVLNITYEKSNNVEYKIKELNLYKTI 1529
QY 1467 QDKLADFPKNNNFVGTADLSTDYNNHNLTKFLSTGMVFENLAKTVLSNLLDGLQGLN 1526
Db 1530 DDKIDLFTKNTDPEAKTKLLNDTTKMDLGLKLLSTGLV-QNFPNTIISKLIEGKFDOMLN 1588
QY 1527 ISOHQVKQCPQNSGCFRHLDBRECKCLLYNKQEGDKCVENPNTCENNINGGCCDADAK 1586
Db 1589 ISOHQVKQCPQNSGCFRHLDBRECKCLLYNKQEGDKCVENPNTCENNINGGCCDADAK 1648
QY 1587 CTEEDSGSKKTKCTCKTDPDYPLEDFGIFCSSSNFLGISFLLILMLILYSFI 1639
Db 1649 CTEEDSGSKKTKCTCKTDPDYPLEDFGIFCSSSNFLGISFLLILMLILYSFI 1701
```

RESULT 7

A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C:Accession: A26868
R:Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium

A:Reference number: A26868; MUID:88011243; PMID:3079521
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TA>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 60.0%; Score 5053; DB 2; Length 1701;
Best Local Similarity 60.1%; Pred. No. 2.9e-149;
Matches 1042; Conservative 225; Mismatches 338; Indels 130; Gaps 30;

QY 1 MKIIFLCFLFIINTQCVTHESYQELVKLEDAVLTGYSLQFQKEMVNEGTS 60
Db 1 MKIIFLCFLFIINTQCVTHESYQELVKLEDAVLTGYSLQFQKEMVNEGTS 60
QY 61 AVTSPGSGKSVASGGSGSVASGGSGSVASGGSGSVASGGSGSRRTNPDSNDS 120
Db 61 AVTSPGSGSVT---SGGSVASVASGG---SGGSVASGGSGSRRTNPDSNDS 114
QY 121 DAKSYADLKHVRPNYLLTIKELKYPOLFDTNHLMTLCNIGHGFKYLDIGEYEinELLYK 180
Db 115 NTQTYADLKHVRQNYLFTIKELKYPOLFDTNHLMTLSKNVDGFKYLDIGEYEinELLYK 174
QY 181 LNFYDOLLRAKLVNDVCANDYCOIPFNLKIRANELDVLKLVFGYRKPDLONIKDNVGKMD 240
Db 175 LNFYDOLLRAKLVNDACANSYCOIPFNLKIRANELDVLKLVFGYRKPDLONIKDNVGKMD 234
QY 241 YKKNKKTITENINELIEESKTTIDKNKNTKBEKKLKYQAOYDLYNKQLEAHNLIS 300
Db 235 YKKNKTTTANINELIEGSKTTIDQKNADNEEGKKLYQAOYDLYNKQLEAHNLIS 294
QY 301 VLEKRDITLKKENIKELDKINEIK---NPPANGNTPTNLLD-KNKKIEEHEKEIK 355
Db 295 VLEKRDITLKKENIKELDKINEIK---NPPANGNTPTNLLD-KNKKIEEHEKEIK 351
QY 356 EIAKTIKFNIDSFTDPLEYLYREKKNINISAKVE--TKRESTPN-EYPNGVYPLS 412
Db 352 EIAKTIKFNIDSFTDPLEYLYREKKNKVDVTPKSQDPTKSVQIPKVPYNGVYPLP 411
QY 413 YNDINNAL---NELNSFGDLINFPDYTKPSKNIYTDN-ERKAFINEIEKIKIEKKI- 467
Db 412 LTDIHNSLAADNKNYSYGLMNP-DTKEKINEXIITDNKERKIFINNKKQIDLEKNIN 470
QY 468 ---ESDKSYEDRSKSLNDITTEYKLLNEIYDSKNNNIDLTNFEKMMGKRYSYKEVL 524
Db 471 HTKEQNKLLEDEYKS---KKDYELLEKEFYEMKNNFNNFDKVDVKIISARTYVNEKQ 526
QY 525 THHTPFASYENSKHNLEKTKALKYMEDYSLRNIVVEKELKYKNIISKIETENIETLVEN 584
Db 527 RYNNKFSNNNSVYVQKLLKALSLEYSLRKGISEKDFNHYYTLTKTLEADIKKLTTEE 586
QY 585 IKKDEQLPEKK---ITKDNKPDKEILEVSDIVKVVQVQVLLMKNKIDELKTKQLILKNV 641
Db 587 IKSSNKKILEKPNKGTHTHSANAS----LEVSQIVLQVQVLLKIDLEKRLKLEFLKNA 642
QY 642 ELKHNTHVNSPKOENKQPPYLLIVLKEIDKLKFWPKVESLINEKKNIKETEGSDNS 701
Db 643 QLKDSLHVPNLYKPNKPEPYLLIVLKEVDKLKEIPKVKMDLKEQAVLSS----- 695
QY 702 EPSTEGETQATTKPGQAGSALGSDSVQAOAQEQKA-----OPVPVVPVPEAKA 754
Db 696 --ITQPLVAASETTEDGGHSTHTL-SQSGETEVEETEETVGTHTTTITLPPKEES 752
QY 755 VTPPPAPVNNKTEN-----VSKLDYLEKLYEFLNTSYICHKYILVSHSTNNEKLLKQ 806
Db 753 APKEVAVWENSTEHKNDNSQALTKTVYKLLKDEFLTSTKSYCHKYLIVSNSSMDQKLEY 812
QY 807 YKLTKEESKLSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFMEIYEKEMVKNLYKL 866
Db 813 YNLTPEENELKSCDPLDLLFNIONNIPVMSFDSLNNSLSOLFMEIYEKEMVKNLYKL 872

Qy	867	KDNDKTNLLEBAKV-----STSVKTLSSS-----MQPLSUTPODKPEVS	908
Db	873	KEENKHLEKEQKQITGTSSTSSPCTVTNTAQASHNSQNOQSNASTQNGVVS	932
Qy	909	AND-----DTSHS-----TNLNSLKLLENILSLGKNKNIOEL-IGOKSSENYEKLKDS	960
Db	933	SGPAVVEESHDPITVLISNDLKGIVLSLNLGNKTKVNPITISTEMEKFEYNILKND	992
Qy	961	TFYNESFTNPFVKADDSINSLNDESKRKKLEEDINKLKKTLQLSFDLYNKYKLERLFD	1020
Db	993	TYFENDDIQFVKSNSKVIITGLT-ETQKNALNDEIKKLDKTLQLSFDLYNKYKLDRLFN	1051
Qy	1021	KKTVCKYKMOIKKLTLLKEQLESKLSLNNPKHYLQNFVSFFNNKKEAEIAETENTLEN	1080
Db	1052	KKLEQODKMQIKKLTLLKEQLESKLSLNNPHNYLQNFVSFFNNKKEAEIAETENTLEN	1111
Qy	1081	TKILLKHVGLVKNYNGESSPLKTLSSESIQTEDNYASLENFKVLSKLEGLKDNLNEK	1140
Db	1112	TKILLKHVGLVKNYNGESSPLKTLSSESIQTEDNYANLEKFRALSKIDGLNDNLHLGK	1171
Qy	1141	KKLSYSSGLHLHIAELKEVINKNKYTGNSPSENNTDVNNALESYKKFLPGTDOVATVS	1200
Db	1172	KKLSYSSGLHLHIAELKEVINKNKYTGNSPENNNKYNEALKSYENFFPE-AKVTTVVT	1230
Qy	1201	E-----SGSDTLEQSOQKPKPASTHVGAESNTITTTSONDVEDDDVIIVIPF	1246
Db	1231	PPQPDVTPSPLSVRVSSGSGSTKEETQIPTSGSLTELOQVVQLONTDEEDDSVLVIPF	1290
Qy	1247	GESEEDYDDLGOVVTGEAVTPSVIDNILSKIENEYEVLYLKLPAQVYRSCLKQLENNVMT	1306
Db	1291	GESEDNEDVLDQVVTGEAISVT-MDNILSGPENEYDVIYKLPAQVYRSCLKQKEKNIIT	1349
Qy	1307	FNVNVKDLLNSRENKENFNKVLSDLPITPYKDLTSSNVVQDPKFKLNKERRDKFLSYN	1366
Db	1350	FNLENLDILNSRLKRRKFLDVLSDLMQFKHISSNEYIIEDSFKLLNSEQKNTLLKSYK	1409
Qy	1367	YIKDSITDITINFANDVGLVGYKILSKYSKLDLSIKKYI-----ND	1406
Db	1410	YIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKVKIKEKEKXFPSPPTPPSPAKTD	1469
Qy	1407	KQGENEKYLPFLNNIETLYKTVNDKIDLFIVHLEAKVLYNITYEKSNVVEVKIKELNYLKI	1466
Db	1470	EOKKESKFLPFLTIETLYNNLVNKKIDDIYLNILKAKINDCNVEXDEAHVYKTKLSDUKAI	1529
Qy	1467	QDKLADFKNNEFVGJADLSTDYNNHLLTKFLSTGMV--FENIAKTVLSNLLDGNLOGM	1524
Db	1530	DDKIDLPKNTNDFEATKKLINDOTYKMDLGLKLLSTGLVQIIFPN---TIISKLEIGKFDQM	1586
Qy	1525	LNISQHCQVKKQCPQNSGCFRHLDEREECKLLNYKQSGDKCVENPNPTCNENNGGCDAD	1584
Db	1587	LNISQHCQVKKQCPENSGCFRHLDEREECKLLNYKQSGDKCBENPNPTCNENNGGCDAD	1646
Qy	1585	AKTEEDSGSGKKITCECTKPDSPYLFDFGIFCSSSNFLGTSFLLIMLILYSFI	1639
Db	1647	ATCTEEDSGSRKKITCECTKPDSPYLFDFGIFCSSSNFLGTSFLLIMLILYSFI	1701

RESULT 8
A45604
major blood-stage surface antigen Pv200 - Plasmodium vivax
C:Species: Plasmodium vivax
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C:Accession: A45604
R:Gibson, H.L.; Tucker, J.E.; Kaslow, D.C.; Kretzli, A.U.; Collins, W.E.; Kiefer, M.C.;
Mol. Biochem. Parasitol. 50, 325-333, 1992
A:Title: Structure and expression of the gene for Pv200, a major blood-stage surface antigen
A:Reference number: A45604; MUID:92158013; PMID:1371329
A:Accession: A45604
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1751 <GIB>
A:Cross-references: GB:M5674; NID:gl60608; PID:g457336
A:Note: sequence extracted from NCBI backbone (NCBIN:83591, NCBIP:83592)

C:Superfamily: major merozoite surface antigen			
C:keywords: surface antigen			
	Query Match	Score 3093.5; DB 2; Length 1751;	
	Best Local Similarity	36.2%; Pred. No. 1.2e-88;	
	Matches 704; Conservative 316; Mismatches 517; Indels 307; Gaps		
Qy	1 MKIIFFLCSPFFIINTQCVTHSYQBLVKKLEDAVLITGYSLFQEKRMVINEGTS	60	
Db	1 MKALLEFSPFIFFTKQCQET-ESYKQLVAKLDKLEALVVDGYELFHKKL----	50	
Qy	61 AVTTSTPGSKGVASGGSGSVASGGSGSVASGGSGSGNSRRTNPSDSSDS	120	
Db	51 -----GENDIKVFETNASANNNN 68		
Qy	121 DAKSYADLKHVRNYYLTIKELKYPQLFDLTNHLMTLCDNIHFKYKLDIGYEINELLYK	180	
Db	69 NOVSV--LTSKIRNLFKSFLELOIPGHTDHLHLIRELAVEPNGIKYLVSEYEFNQLMHV	126	
Qy	181 LNFYDILLRAKLDVANDCYQIPFNLUKIRANELDVLUKLIVFYGRPLDNKONVKMED	240	
Db	127 INFHYDILLRAKLDHMCADHYCKPHEHLKISDKELMDMLKVLVGYRKPDLNKKIDIGKLET	186	
Qy	241 YIKNKKTIEINEL-TEESKK-----	261	
Db	187 FITKNKITIKNISDLIIAENKKRSGHPTTTNGAGTOPANGSIAAASSETQISGSSNSG	246	
Qy	262 -----TIDKNKNATKEEEKKLYQAQYDLSIYNKQ 291		
Db	247 SSSTGSSNSGSSSTGSGTGSGQPAPAAASSTNANYEAKIYQAVYVITFYTNQ	306	
Qy	292 LEEAHLISVLEKRIDLTKKNENIKELDKINEIKNPPA---NSGTPMTLLDKNKKIE	348	
Db	307 LQEAQKLIATVLEKRVKLVKHKDKIVLQEOVAKEKELPSDYPTNTNLTNVHRAESKIA	366	
Qy	349 EHEKEIEIAIKFNIDSLSFTDPLEYLYLREKNKI-DISAKVETKESTEPNE-----	402	
Db	367 ELEKKIEIAIKTVNFDLDGLFTDAEEYLYLREKAMAGTLIIPESTKSAGTPEKTVPTL	426	
Qy	403 ---YPNGVTPYPLSYNDINNALNELNS---FGDLINPFDYTKEPSKNI-YTDNRKRFKINE	455	
Db	427 KETPHGISVALAENSIVELIEKIGSDETFGDLQNP-DDCKQPKGLLINETRKLELLEK	485	
Qy	456 IKEXIKIEKKKIESDKSKSIEDRSKSLNDITKEYEKLNEIYDSKFNNNIDLTFEKMKG	515	
Db	486 IMNKIKIEEDKLPNLKYEYBEKYEAKVNEFKPAPNFYEARDLNTLVENKFDDFKKK	545	
Qy	516 RYSYKVKVLTHTNFTASYENSKHNLEKLTALKYIMEDYSLRNIVVEKELKYNNLSKIE	575	
Db	546 REAYMEEEKKLES--CSYEQNSNLINKLKQLYLEDYVLKRDIAODEIKHFSFMEWKLK	603	
Qy	576 NEIETLVENIKKDEQLFEKKITKDKNKPDKEKILEVSDIVKVOQVKVLLMNKIDELKKTQ	635	
Db	604 SEIYDLAQEIRKNENKL-----TIENK-----PDFSGVVELQVKVLIKKLEALKNVQ	652	
Qy	636 LILANVELKHNIHPNSYKQENKQEPYYLILVKLEIKDKLVFMPKVESLINEEKK-----	690	
Db	653 NLLKANAKVDLLYIPKYVKTSEKPEYYLMLVKREIDKDKDFPIKTESMIATEKNKPTVA	712	
Qy	691 --NKTGQS--DNSPSTG-----EITGOATTKPG-----QAQSA-LE 726		
Db	713 AADTVARGQSLRGASSETGTTGNTVNAQTAVYVQPHQVNVNAVTVQPGTTHQAQGAETQ	772	
Qy	727 GDSVOAQAEQKQA-----OPVPVPV-PE-AKAQVPTPPAPVNNKTEVNSK 771		
Db	773 TNSVOAAQVQVOTPAGAGGVASQTTSQAPAPTAQSPAPAPPTSPAPAAVAPATMSK	832	
Qy	772 LDYLEKLYEFNTSYICHKYLIVSHSTMENKILKQYKITKEESKL--SSCDPLDLFFNI	829	
Db	833 LEYLEKLLDLFLKSAYACHKHIFVNTSTMKKELLDQYKLNADQNKINETKCDLDDLLFNW	892	
Qy	830 QNNIPVMYSMPDLSNLSOLFMEIYEEKVMCNVLKLDND-KIKNLLFEAK-----VS	893	

Qy	1420	N I E T L Y T V N D K I D L F V I H L E A K V L N Y T Y E K S N E V E V K I E L N Y L K T I O D K L A D F K K N N F	1479
Db	1517	S L Q K E Y E S L S K V N T Y T D N L K K V I N N K Q L E A E A T T Y V K K I Q D Y N K M D E K L E E Y K	1572
Qy	1480	V G I A D L S T D Y N H N N L L T K F L S T C M V F E N L A K T V L S N L L D G N L O G M L N I S Q H C V K K O C P Q	1539
Db	1573	---S E K N E V K S G L L E K L M K S K L I K E N E S K E I L S O L L N V Q T L L T M S S E H T C I D T N V P D	1629
Qy	1540	N S G C F R H L D R E E C K L L N Y K O E B D C K V P N P T C N N N G G C D A D A K T C E D S G S N G K K I	1599
Db	1630	N A A C Y R Y L D G W E R C L L F E K E G G C V P G S N V T C K D N N G C A P E A E K W M T D S	1685
Qy	1600	T C E C T K P D S V L P D G I F C S S N F L G I S F L L I M L I L	1635
Db	1686	V C K T K E G S E P L F E G V F C S S S F L S F L L M L L F L	1721
RESULT 11			
S06286			
major merozoite surface antigen precursor - malaria parasite (Plasmodium fa			
N:Alternate names: 190K protein; polymorphic schizont antigen			
C:Species: Plasmodium falciparum			
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jun-2000			
C:Accession: S06286			
R:Certa, U.; Rotmann, D.; Matile, H.; Reber-Liske, R.			
EMBO J. 6, 4137-4142, 1987			
A:Title: A naturally occurring gene encoding the major surface antigen prec			
A:Reference number: S06286; MUID:88166657; PMID:3327688			
A:Accession: S06286			
A>Status: not compared with conceptual translation			
A:Molecule type: DNA			
A:Residues: 1-1060 <CER>			
C:Superfamily: major merozoite surface antigen			
C:Keywords: surface antigen			
Query Match 33.6%; Score 2829.5; DB 2; Length 1060;			
Best Local Similarity 55.5%; Pred. No. 1e-80;			
Matches 609; Conservative 145; Mismatches 255; Indels 89; Gaps			
Qy	1	M K I I F F C S L F F I I N T Q C V T H S Y Q E L V K K L E A L E D A V L T G Y S L F Q K E K V L N E G T S G T	60
Db	1	M K I I F F C S L F F I I N T Q C V T H S Y Q E L V K K L E A L E D A V L T G Y S L F Q K E K V L N D G A N T Q	60
Qy	61	A V T T S T P G S K G S V A S G G S G S V A S G S V A S G S G S V A S G S G S R R T N P S D N S S D S	120
Db	61	V W A K P A D A V S T Q S A K N P P G A T V P S ----- C T A S T K G A I R S P G A ----- A N P S D D S S D S	108
Qy	121	D A K S Y A D L K R V R N Y L L T K E L Y P O L F D L T N H M L T L C D N I H G F K Y L I D G V E E I N E L L Y K	180
Db	109	D A K S Y A D L K R V Q N Y L F T I K E L Y P E L F D L T N H M L T L C D N I H G F K Y L I D G V E E I N E L L Y K	168
Qy	181	L N F Y F D L L R A K L D V C A N D Y C Q I P P N L K T R A N E L D V L K K L V F G Y R K P L D N I K D N V K M E D	240
Db	169	L N F Y F D L L R A K L D V C A N D Y C Q I P P N L K T R A N E L D V L K K L V F G Y R K P L D F I K D N V K M E D	228
Qy	241	Y I K N K K T I E N I N E L I E S K K T I D K N K A T E E E K K L Y Q A Y D L S I N K O L E E A H N L I S	300
Db	229	Y I K N K T T I A N I N E L I E G S K K T I D K N K A D N E E G K K L Y Q A Y D L F I N K O L Q E A H N L I S	288
Qy	301	V L E K R I D T L K K N E I K E L D K I N E K N P P A N S G W N T P N T L D K N K K I E E H E K E I K E T A K T	360
Db	289	V L E K R I D T L K K N E I K K L E D I D K I D A E K P T T G V N O I L S L R L E K E S R H E E K I K E T A K T	348
Qy	361	I K F N I D S L F T D P L E Y L Y R E K N K N I D I S A K V E --- T K E S T E P N - E Y P N G V T Y P L S Y N D I N	417
Db	349	I K F N I D R L F T D P L E Y L Y R E K N K K V D V T P S Q D P T K S V Q I P K V P Y P I N G V I Y P L P L D I H	408
Qy	418	N A L --- N E L S F G D L N P F D Y T K E P - S K N I Y T D N - E R K K F I N E I K E K I K I E K K I --- E	468
Db	409	N S L A A D N K S Y G D L M N P - H Y F K E I N E K I I T D N K E R I F I N T K K I Q D L E E K I N I T K E	466
Qy	469	S D K S Y E D R S K S L I D T K Y E K L L N E I Y D S K F P N N I D T N F E K M G R Y S Y K V E K L T H H N	528


```
QY 1 MKIIFLCSEFLFIINTOCVTHESYQELVKKLEAEDAVLTCGYSLFQEKWVLNBSGTSGT 60
Db 1 MKVIGLLSFVFAIKCKSETTEVNDLIQKLEKLESISVGLGLEFQKQSVIIN----- 54
QY 61 AVTTSTPGSGSVASGGSGSVASGSGSVASGSGSVASGSGSGNSRRTPNPSDSSDS 120
Db 55 -----ATOPTETIDPF 65
QY 121 DAKSVADIKLRVRYNLLIKELKYPQLFDLNNHMLTLCDNTHGFKYLDIGVEEINELLYK 180
Db 66 TNHNFQ--QQQDQVTFTEGLFTQETELVNLKALPNRYGVYGLIESKEEFENGLMHA 122
QY 181 LNFYDLRLAKLNDVANDYCOIPNKLIRANELDVLAKLVGFKYRKLNDIKDNVGMKED 240
Db 123 INFYDVLRLKNDWCANNYCEIPEHLKISEBETEMLAKVILGFKPIENIQDDLEKLEI 182
QY 241 YIKNNKKTENINELIESKTTI-----DNKNKATKEEEKKKLYQAOYDLSIYNKQ 291
Db 183 YIERNKETVAALNATIAEETKKIQPEGNECDNDASCDSDKYNKKKPIYOAMVNYFYKKQ 242
QY 292 LEEANLISVLEKRTDITLKKNIENIKELLDKINEIKNP-----PANSQ 334
Db 243 LAEIQKVVEVLEKRVSTLKNDAIKPLWQOIEVLNAAVPVTAETQIVTGGOSSTSPGSGG 302
QY 335 N-----TPN-----TLLDNK 344
Db 303 SSASCTSSGQASAGTGVQEAQNTVASVTVTFVSGONGEASNPQTAQVOPVPTLTLEKQ 362
QY 345 KKIIEHEKEIKETIAKTIKFNIDSFTDPLELEYLYUREKN-----NIDISA--KVETKESTE 399
Db 363 KKIAGLYAQIAKEIAKTIKFNLEIGFVDPTELEYFKKKEKESCNLSTSSCKKNKASETII 422
QY 400 P--NEYPNGVYPLSYNDI-----NNALNELNFSQDLINPDYTKPEKNIYTDNERKKF 452
Db 423 PUTIRYPNGISYPLPENDVYNNKIANNAB--TTYGDLTHP--DNTPLTGDLATNEQARLD 479
QY 453 INEIKERIKIKTESDKSKSLNDITTKYEKLLNBEIYDSKNNNIDITNFEMK 512
Db 480 IKAIKKIKAEKLETKLTNTYDNKLTFTFNQOKTPFEKAKEAFYESKRNKLTSEIFEK 539
QY 513 MKKRYSVKVEKLTHTNTASTENSXHNLEKLTALKUMEDYSLRNVVYKEKLYKKNLIS 572
Db 540 KTKRDEYTKK--TELNT--CEYGNTRKELINKQLNQLQDYSLRDKDIISNEIEYFESNKK 597
QY 573 KIENIEIETLVENIKKDEQLFEKTKTKDENKPEKILEVSDIVKVQOKVLLMKNKIDEL 632
Db 598 ELQYINIRLAEVQAKQNVLA-----SKDVPLSTLVELQIQKSLLTQKIOELN 646
QY 633 KTOILKLVKELKHNTHVPSNYKQENKQEPYLLIVLKKKEIDKLVKPMKPVESLINEKNI 692
Db 647 KTEVSLNKAQLKDLVVPKTYGNECKPEPYLLIAVKKEVDRLAQPIPKIESMIAKERM 706
QY 693 K-----TEGQSDNSEPSTEGEITGQATTPGQOAGSA-----LEG 727
Db 707 EQGPAITGESEVPSPGSAESTDSTQSTSSSSSSSTPAAAESSATLPEAPAPAEA 766
QY 728 DSVQAOAQOQAOQAPVPVPEAKAQVPTPAPVNNKTENVSKLDYLEKLYEFLNTSYI 787
Db 767 ASPSTEAEBEETIPTTQETQPSQAASSTTFAKPV-----MTKLYYLEKLOKFLVFSYS 820
QY 788 CHKYLIVSHSTMNEKILKQYKITTKEES--KLSSCDPLDLLFNIGNNPVPMVSMFDSLNN 845
Db 821 CHKVYLLQNSNTINKDALSXYALTSEDKIRTLKRCELDVLLAIONNNPTMYSLYESIVD 880
QY 846 SLSQLFMEIYEKEMVCNLYKLD--NDKIKNLEEA----- 879
Db 881 GLQNIYTELEYEKEMHYIYKLDENPSIKSLVKAGVIEPEPVAAPTPVTPAATBQOQOQ 940
QY 880 --KKVSTSVKTLSSSSMQPLSLITPODKPEVSANDTSHS----- 916
Db 941 ATPDVQSDAPAPSDVSOQPETPVTTTPEVTTSTSEASSAPGEGTPSGAGASGTGATA 1000
```

```
QY 917 -----TNLNNSLKLFENILS-----LGKNKNIYQELI 943
Db 1001 SNAATPAGTSAGSASNASNTSDVTPPAAAAAAVSTSTPAPAAQPAANSQSGNPDGSR 1060
QY 944 GOKSSE-----NFYEKIKDSDTFYNESTNFVKSADDINSUNDSKRKKLEB 992
Db 1061 SRAESEEDMPADDFELDNLYKSLQOIDG--NNTFEINFKSKKELIKALTPK--KVNLQYL 1118
QY 993 DINKLKKTKLOISFDLYNKYKLLERLFDKKKTVGKYKQIKKLTLLKKEQESKLNSLNP 1052
Db 1119 EIAHLKLESEHYDYRSTYKLERLYNKHQIQTNQIRDLISILKARLKRKQTLNGV 1178
QY 1053 KHVLFNSVFFNKKKEAETAETENTLENTKILLKHYKGLVKKYNGESSPLKTLSESIO 1112
Db 1179 FYILNGYVNFNKRREAEKQYVDNALKNTDMLLKYKARTKYFTSEAVPLKTLKASLDR 1238
QY 1113 EDNYASLENFKVLSLEKGLKDNLEKKLSYLSGSLHLLIAELKEVITKNKNYNSPS 1172
Db 1239 ESNYLKIEKFRAYSRLERLRLKKNINLGRISYVSGLHHVPEEFKELIKDKDYTGKKNP 1298
QY 1173 ENNTDVNNALESYKFLPEGTDTVAVVSESGDITLEQSQPKKPASTHYGAESNTI----- 1227
Db 1299 DNAEPTNAFEQYKELLPGKVTST--PAVAVTTTLLAADAPATPEGAVPGAVPGAVP 1357
QY 1228 -----TTSQNVDDVDDVIIPTFGSESEEDYDGLQGVVTGEAVTPSVIDNLSK 1276
Db 1358 GAVPGAVPGSGTDTTRVAGSSVDD-----NED--DDIYQIASQSEDAPEKD--ILSE 1405
QY 1277 IENEYEVLYLKPAGVYSLKKQLENNVWTFVNVVKDILNSFRNKRNFKNVLESDLIPY 1336
Db 1406 FTNESLYVTVTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNKKSKQRNDPFLVLSHIELDLF 1465
QY 1337 KDLTSSNVVDPYKFLANKEKDFELSSVNYIKDSITDIDFANDVGLVGYTLSSEKYSK 1396
Db 1466 KDLSTNNKVIIRNPYQLLNDKKDKQIVNLKYATKGINEDIETTTGDFKFNKMWELYNQ 1525
QY 1397 LDSIKKYI-----ND--KQGENEKYLPFLNNIETLYKTVNDKIDFLVHLEAKVLYNTY 1448
Db 1526 LAAVEQIATIBAEATNDTNKEKKYIPILEDLKLGYETVIGQAEYSBELQNRDLYKN 1585
QY 1449 EKSNEVEVIKELNYLKTITQDKLADF-----KKNNFVFGIADLSTDYNNHNLTKFLSTGMV 1504
Db 1586 EKAEPFILTNLKELYIQIDEKLDEFVEHAENKNHASTA-----LNNLKNKSLV 1634
QY 1505 PENLAKTVLSNLLDGNLQGM--LNIS--OHQV--KQCPONSCFRRHLDRECKCLLYNK 1560
Db 1635 GEGESKKILAKML--NMDGMLLGYDPKHVCYVDTRDIPKNACGFRDDNGTEBWRCLLYGK 1692
QY 1561 Q--EGDKCVENPNPTCNENNGGCDADAKTEEDSGSNGKKITCECTKPDSPYPLFDGIFCSS 1619
Db 1693 KGEGTCTVENNPTCDINNGCDPTASQNAESTENSKKIICTCKEPTPNAYVEGVFCSS 1752
QY 1620 SNFLGISFLILLMLILYS 1637
Db 1753 SSFMLGSLILLIITLIVFN 1770
```

RESULT 13

A45546

major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi

C:Species: Plasmodium chabaudi chabaudi

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999

C:Accession: A45546

R:Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca

Mol. Biochem. Parasitol. 43, 231-244, 1990

A:Title: Molecular cloning and sequence analysis of the gene encoding the major meroz

A:Reference number: A45546; MUID:91218805; PMID:2090945

A:Accession: A45546

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1785

A:Cross-references: GB:M34947; NID:g160597; PID:g160598

C:Superfamily: major merozoite surface antigen

C;Keywords: surface antigen

Query Match 29.8%; Score 2509; DB.2; Length 1785;
Best Local Similarity 31.6%; Pred. No. 1.5e-70;
Matches 598; Conservative 324; Mismatches 605; Indels 366; Gaps 35;

Qy	1	MKIIFLC	SFLFI	INTQCV	THSY	QELV	YKKL	EADVL	TV	YSLF	FQK	EMV	LN	BEG	TST	60								
Db	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1								
Db	1	MKAIGL	FSVFA	YCKSET	IGVY	NDLV	HLEK	LELS	VEGL	ELFK	QSVI	YNAQ	SPET	60										
Qy	61	AVTIST	PGSG	VASGG	SGSV	ASGGS	VASGG	SGSV	SGSV	SGSV	SGSV	SGSV	SGSV	SGSV	SGSV	120								
Db	61	PVDPF														68								
Qy	121	DAKSYAD	LKHR	VNYL	TIKEL	YPO	LDTN	HMLT	LCNI	HGFK	YLID	GCYE	INELL	YK	180									
Db	69	----	EFAQ	KLPF	ILKPE	LGFTEQ	TEL	VNLK	ITLGP	NKYGLK	YLIESK	EENEL	MHA	122										
Qy	181	LNFYD	LLRAK	LNDV	CANDY	CQ	PNFLK	IRAN	ELD	VKLK	LVFGY	RPLON	IDNVG	KMED	240									
Db	123	INFYD	YLRD	LDMC	ANNY	CEI	PEHLK	INVE	IEMLK	KVLV	GYR	PINI	ODDL	KLEE	182									
Qy	241	YIKNK	KTIT	INEL	BESK	TT	----	----	----	----	----	----	----	----	291									
Db	183	YIARK	KATA	EYTL	NLTITE	TKK	TT	PEET	DCN	TD	CNTKY	GKKAI	YQAM	NI	VPYKKO	242								
Qy	292	LEE	AHNL	SIVLE	KRID	PLK	KNENI	KELL	DR	IN	EIKN	PPAN	SG	----	334									
Db	243	LAEIK	KVIE	LEK	RVAL	TAK	NEAL	KPLQ	QIEA	IR	GP	PAVE	TQ	EG	ISSE	ETKONSTE	302							
Qy	335	--	NTPN	TLLD	----	----	----	----	----	----	----	----	----	----	342									
Db	303	SSNTK	TTTT	OKAV	TTOT	ATKAT	GTNT	GTNT	GTNT	GTNT	GTNT	GTNT	GTNT	GTNT	GTNT	362								
Qy	343	----	----	----	----	----	----	----	----	----	----	----	----	----	386									
Db	363	EPVQ	VPAQ	VL	TBE	EAKK	TAEL	YAQI	KE	LAKT	IKEN	LDGIF	VDP	VE	LEYE	YKKEKNESC	422							
Qy	387	DISA	----	----	----	----	----	----	----	----	----	----	----	----	437									
Db	423	HST	SSCHK	NTPT	VTI	PLN	VRYP	NGIS	YPLTE	EWY	SKIA	HAAAE	----	----	480									
Qy	438	EPSK	NIYD	NERK	FE	INEI	KE	TIK	IEKK	KTES	OK	SYED	RS	LSND	IT	KEYEKL	NEIYD	497						
Db	481	ITED	LTTNE	QARN	LKAI	KKK	LEAE	Q	LEU	KDDY	DYTK	LAA	FNG	QKTP	FE	AKA	KFFE	540						
Qy	498	SKFN	NIDL	NFE	KMG	KRY	SYK	VEK	LTH	HTNT	PAS	YENS	KN	LEK	LTK	KALK	MYED	SLRN	557					
Db	541	SKPR	NLT	TD	IFDD	FKTK	REY	NNKAA	----	----	----	----	----	----	----	----	----	598						
Qy	558	IVVE	KEL	YK	YN	LISK	TIEN	IE	ITL	VENI	KDE	QL	FEK	IT	DEN	KPD	BE	IL	VS	DIVK	617			
Db	599	EIVN	TEY	FSN	KSEL	QY	WIN	LANA	VQA	KNIL	VA	----	----	----	----	----	----	647						
Qy	618	QVQ	VLLM	NK	IDEL	KK	TOLL	K	NV	ELK	HNTH	VPNS	V	KQEN	KQ	PPY	YLIV	LKE	ID	KLVF	677			
Db	648	QIQ	SLLT	KLIB	EOLN	TE	FS	LN	KAHL	KDKI	YV	PQT	Y	GK	EK	PPY	YLIA	IK	EID	RLAF	707			
Qy	678	MPK	V	S	L	I	N	E	K	K	N	I	K	T	E	G	----	----	----	----	721			
Db	708	IPK	ID	MI	E	K	E	K	O	E	H	V	A	T	E	G	S	A	S	A	S	767		
Qy	722	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	757			
Db	768	G	T	E	T	E	T	P	A	A	K	P	A	G	A	S	T	G	A	T	T	827		
Qy	758	P	A	P	V	N	K	T	E	N	S	K	L	D	L	E	K	L	E	F	L	N	815	
Db	828	P	A	P	A	A	P	A	K	P	W	M	K	L	E	K	L	E	F	L	N	887		
Qy	816	K	L																					

Qy	875	LL-----	-----BEAKKVSTSVKT--	888	
Qy	876				
Db	948	LLVKAGVIDPEVPAPETAPETAPETAPETAQEAPOQESAQAPEAAETTP	1007		
Qy	889	LSSSSQPLSLTPQDKP-----	-----EVSANDDTSH	915	
Qy	890				
Db	1008	AESAESTPTPKATATPTSETVTQEGTTPAAPKAQEGASSAPAPAKPAPATVGTQ	1067		
Qy	916	STNLNLSKLFIENILSGRNKNYQBELIGQKSSSENFYEKILKSDTFYNESFTNFVSKSA	975		
Qy	917				
Db	1068	STNVEGSTQV-----	-----RAESEDEFVDDFVDNFYKSYLQQVDG--NNTQDFIDFKSKK	1118	
Qy	976	DDINSLNDESKRKKLEEDINKLKKTLQLSFDLYNXYKLLKRLPDKPKKTVGKYKMQIKKL	1035		
Qy	977				
Db	1119	ELINALTPE-KVNLQYLDIAHLKESEHYNRYKYKLERLYQKHQJQIEAANQKVEI	1177		
Qy	1036	TLIKEOLSKLNLNPKHVLQNFVSFFFNKKKEAETAETENFLENTKILLKHYKGLVKY	1095		
Qy	1037				
Db	1178	SVLSRLLRKKYINGTFYVLSGFANFFNRREAEKQYVDNAIKWTMLLAYYKARSYF	1237		
Qy	1096	NGSSPLKTLUSESIOTENYASLENFKVLSKLEGLKDNLNLEKKLSYLSUGLHLHJIA	1155		
Qy	1097				
Db	1238	TSEAVPLKTLTKTSIDREANYLKIEKFRAYSLELRLLKKNLNKGERITYYVSGGLHHVFE	1297		
Qy	1156	ELKEVINKNWTCNSPSENTOVNNALESYKFLPEGTD----	-----VATVVSSESGSDTLEQS	1210	
Qy	1157				
Db	1298	EFKELLKNKGTGKTNPENAPEVIRAKFQYKELLPGATTPAPVVPVPAAPATAAPAA	1357		
Qy	1211	QPKKP-----	-----ASTHVGAESNTITTTSONYDDEVDVVIIPIFGESEEDYDDLQ	1258	
Qy	1212				
Db	1358	DAPVPAASAAAAASGSGSAATTEGEAATTVVASSDND-----	-----DDDDDDMQ	1402	
Qy	1259	VTGEAVTPSVINDILNLSKINENEYVLKPLAGVYPSLKKOLENNVNTFNVNVKDLNSR	1318		
Qy	1403	IANAQS--TDBEVDKIIDAKFKSENEYIYTKSLGNTYKSEKHHMLKBEFSMIKEDIMTGLNYK	1461		
Qy	1319	FNKRENFKNVLESOLPYKDLTSSNVKDPYKFLNKKEROKFKLSSYNTIKDSIDTDINF	1378		
Qy	1462	LEKRNDFDLVSLYELAFKXIDINTNFVKVPYQLLNDKKOKQMINLKYAIKGVTEDET	1521		
Qy	1379	ANDVLGYKILSKSKYSDDLSDKKIY-----	-----NDKOGENEKYLFPFLNNTIETLYKTVNKKID	1433	
Qy	1522	ATDGIIEFFNMKIELYKPOLNAVNEQFAATIGTEPTDAEKKKXAPIEDULKGLYEITLNGAE	1581		
Qy	1434	LFVHLEAKVLNITYEKSNNVYKIRNELNLYKTIOQLADF----	-----KKNNVFVGIADLSTDY	1489	
Qy	1582	EFSELQHLKENYKIEKAGFDILMANLEYIRIDEKLEDFVESAEKKNHIASA-----	-----	1635	
Qy	1490	NHNNLTFKLSTGWPFENIAKTVLSMLLDGNLOGMLNI--	-----SQHCYVKKQOPONSGCFRHL	1548	
Qy	1636	-----LNNLNKSLGVTEGESKKILAKMLNMDAMDLLIGISNHVCI	-----ISTSTPDNAGCFRYDD	1690	
Qy	1549	ERECECKLLNYKO--	-----RGDKCVENPNPTCENNNGGDAADAKTCEEDSGSN--	-----GKKITCECT	1604
Qy	1691	GTEWRCLLGTFRKDDGNRCVADDAPVCNNNGGCDKNADCREVENTORDPSKKIVCTCK		1750	
Qy	1605	KPDSYPLFDGIFCSCSNFLGIFLILMILILYS	1637		
Qy	1751	EPNPNAYIAGVFCSSGFMGLSILLITLITLIVEN	1783		

RESULT 14
S47282

merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71)
 C:Species: Plasmodium falciparum
 A:Variety: strain RO-71
 C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
 C:Accession: S47282
 R:Tolle, R.; Bujard, H.; Cooper, J.A.
 submitted to the EMBL Data Library, July 1994
 A:Description: Plasmodium falciparum: recombination within the C-terminal region of m
 A:Reference number: S47282

A:Accession: S47282

A:Molecule type: DNA

A:Residues: 1-651 <TO>

A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258

A:Experimental source: strain RO-71

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 25.5%; Score 2146; DB 2; Length 651;
Best Local Similarity 64.5%; Pred. No. 8.3e-60;
Matches 422; Conservative 82; Mismatches 114; Indels 36; Gaps 5;

QY 1019 FDKKTVGKYKMQIKKLLKQLSKLSNLPKHVQLNFSVFNKKKEAIAETNTL 1078

Db 1 FNKKELGQDKMQIKKLLKQLSKLSNLPKHVQLNFSVFNKKKEAIAETNTL 60

QY 1079 ENTKILLHKVGLVYNGESSPLKTLSEESTQIEDNTVASLENFKVLSKLEKLDNLN 1138

Db 61 ENTKILLHKVGLVYNGESSPLKTLSEESTQIEDNTVANLEKFKALSKIDKGLNDNLHL 120

QY 1139 EKKKLSYSSGLHLIAELKEVIKKNYKNTGNSPSENNTDNNALSEYKFKFLPEGTDVATV 1198

Db 121 GKKLSFSSGLHLITELKEVIKKNYKNTGNSPSENKKVNEALASYENFLPE-AKVTV 179

QY 1199 VSE-----SGSDTLEOSQPKKPASTHVGAESNTITTSQNVDDVDVVIIVP 1244

Db 180 VTPPPQDVTPSPLSVRVSGSGSTKEETQIPTSGSLLTTELQOVVOLQNYDEDDSLVVL 239

QY 1245 IFGESEEDYDGLGVVVTGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNV 1304

Db 240 IFGESEEDNDEYDQVVTGEAISVT-MDNILSGFENEYDVIYLDKPLAGVYRSLLKQIEKNI 298

QY 1305 MTFNPNVADKILNSRFNKNFNKLESLLPYKDLTSSNVVVKDPYKFLNKEKDKFLSS 1364

Db 299 FTFENLNDILNSRLKKRYFLDLESOLMQPKHSSNEYIIEEDSKLNSQKNTLKS 358

QY 1365 YNVIKDSIDTDINFANDVLGYKIUSEKYKSDLSIKYI-----N 1405

Db 359 YKIKESVENDIKFAQEGISYEKYLAKYKDDLESIKKVIKEKEFPSPPTTPSPART 418

QY 1406 DRQGENEYLPPLNIETLYKTVDKIDLVTHLEAKVLYNTEKSNVEVTKELNLYKT 1465

Db 419 DQKRESFPLPTNIETLYNNLVKIDYLLNLAKINDCNVEKDEAHVKITKLSDLKA 478

QY 1466 IQDLADFKNNFNFGIADLSDYNNHLLTKFLSTGMVFENLAKTVLSNLDNGLQML 1525

Db 479 IDDKIDLFKNHNDFAIKKLLINDTKDKMLGKLLSTGLV-QNFPNTIISKLEKGFQDML 537

QY 1526 NISQHCYKVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADA 1585

Db 538 NISQHCYKVKQCPQNSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADA 597

QY 1586 KCTEEDSGNGKKTCTCKPDSYPLFDGIFCSSSNFLGISFLLLMLILYSFI 1639

Db 598 KCTEEDSGNGKKTCTCKPDSYPLFDGIFCSSSNFLGISFLLLMLILYSFI 651

RESULT 15

A45545

major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)

C:Species: Plasmodium falciparum

C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000

C:Accession: A45545

R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.

Mol. Biochem. Parasitol. 49, 29-33, 1991

A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1

A:Reference number: A45545; MUID:92131048; PMID:1775158

A:Accession: A45545

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-400 <BLA>

A>Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBI:77621)

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 24.9%; Score 2101; DB 2; Length 400;

Best Local Similarity 100.0%; Pred. No. 1.2e-58;

Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1240 VIVPIFGESEEDYDGLGVVVTGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQ 1299

Db 1 VIVPIFGESEEDYDGLGVVVTGEAVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQ 60

QY 1300 LENNVMTNVNVKDTILNSRFNKNFNKLVESDLTPYKDLTSSNVVVKDPYKFLNKEKRD 1359

Db 61 LENNVMTNVNVKDTILNSRFNKNFNKLVESDLTPYKDLTSSNVVVKDPYKFLNKEKRD 120

QY 1360 KFLSSYNYIKOSIDTDINFANDVLGYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFN 1419

Db 121 KFLSSYNYIKOSIDTDINFANDVLGYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFN 180

QY 1420 NIETLYKTVDKIDLVTHLEAKVLYNTEKSNVEVTKELNLYKTIOKDLADFKNNNF 1479

Db 181 NIETLYKTVDKIDLVTHLEAKVLYNTEKSNVEVTKELNLYKTIOKDLADFKNNNF 240

QY 1480 VGIADLSDYNNHLLTKFLSTGMVFENLAKTVLSNLDNGLQMLNISQHCYKVKQCPQ 1539

Db 241 VGIADLSDYNNHLLTKFLSTGMVFENLAKTVLSNLDNGLQMLNISQHCYKVKQCPQ 300

QY 1540 NSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADAKTEEDSGSGNGKI 1599

Db 301 NSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADAKTEEDSGSGNGKI 360

QY 1600 TCECTKPDSPYPLFDGIFCSSSNFLGISFLLLMLILYSFI 1639

Db 361 TCECTKPDSPYPLFDGIFCSSSNFLGISFLLLMLILYSFI 400

RESULT 16

A25814

glycoprotein 185 - malaria parasite (Plasmodium falciparum) (isolate Honduras I)

C:Species: Plasmodium falciparum

C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 09-Jun-2000

C:Accession: A25814

R:Howard, R.F.; Ardeshir, F.; Reese, R.T.

Gene 46, 197-205, 1986

A:Title: Conservation and antigenicity of N-terminal sequences of GP185 from differen

A:Reference number: A25814; MUID:87106855; PMID:3542719

A:Accession: A25814

A:Molecule type: mRNA

A:Residues: 1-196, 'T', 198-233 <HOW>

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein

Query Match 14.1%; Score 1187; DB 2; Length 233;

Best Local Similarity 95.1%; Pred. No. 1.2e-30;

Matches 232; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 1 MKIIFFLCFLFFIINTQCVTHESYQELVKKLEALEDAVLGTYSLFQEKMYLNEGTSGT 60

Db 1 MKIIFFLCFLFFIINTQCVTHESYQELVKKLEALEDAVLGTYSLFQEKMYLNEGTSGT 60

QY 61 AVTTSTPGSKSVASGGSGGSGSVASGSGSVASGSGSVASGSGSNRRTPNSDSSDS 120

Db 61 AVTTSTPGSKSVASGGSGGSGSVASGSGSVASGSGSVASGSGSNRRTPNSDSSDS 108

QY 121 DAKSYADLKHVRNVLTKELKYPQLPDLTNHMLTLCDNHIGFKYLDGYEETNELLYK 180

Db 109 DAKSYADLKHVRNVLTKELKYPQLPDLTNHMLTLCDNHIGFKYLDGYEETNELLYK 168

QY 181 LNFYFDLLRAKLDVNCANDYCOIPFNKIRANELDVLKLVFGYRKPLDNTKDNVGMED 240

Db 169 LNFYFDLLRAKLDVNCANDYCOIPFNKIRANELDVLKLVFGYRKPLDNTKDNVGMED 228

QY 241 YIKK 244

|||||

[illegible]

Qy	729	SVQAQAEQKQAOQPPVPVPPEAKAQV-----PTPPAPVNNK-----	-----TENV	769
Qy	730	SVQAQAEQKQAOQPPVPVPPEAKAQV-----PTPPAPVNNK-----	-----TENV	769
Db	960	HKEMVAELERKRAHDLAVLVEEQHKAIIKLGEEHKVYVAGIEEKYAEIAKLAEBHKDVV	1019	
Qy	770	SKL-----DYLEKLYEFLNVSYCHKIILVSHSTMNKIL-----KQYKTKKEESKL	817	
Db	1020	TKLGEQHKIEAKLED-----GHKEVVNEVKKNASLLNKLNEENHKNEIMKILKEEHE	1072	
Qy	818	SSCDPDLLE-----NTONNIPVMYSMFDLSNLSQLPMWEIYEKVMCVNLY-----	864	
Db	1073	SASDLVEKLYQKDEEVKYNKNRIIEELTNVIKDLNDSIMCYKKQILEEYKRENEYNEEINK	1132	
Qy	865	-----KLKD-NDK-----IKNL-----LBEAKVSVSVKTLSSSSM-----OPLSLT	900	
Db	1133	LKIVONEMKDMDKKLLEKENELKLNKKLSNYKVFETKENTYKNSVMYNNENKRIIVD	1192	
Qy	901	POOKPEVSANDDTSHTNLSNLSKLFE-----NILSLGKNKNIYQELI-----	943	
Db	1193	SVCKENISBDEVBGKGKMLKMTLSLKKKERNFISINDNKNSESSBELVDITKSAYINKIEM	1252	
Qy	944	-----GOKSENPFYKILKDSOTYN--ESFTNPFYKSRADOI-----NSLNDSEKRX	988	
Db	1253	KKEIEDNGKRIEDLNKRIIDLNELINDENMKVITDGNLKKKEIEIKONKLNKEKNE	1312	
Qy	989	-----KLEEDINKLK-----TLOLSFDL--YNK--YKLKLERLEDFKKKTV	1025	
Db	1313	NTEILNLNDDOIKLKKEISEWKDEEKLTKENIKLKNDIEQINKEYKIKBENL-----M	1366	
Qy	1026	GKYMQIKKTLTKLEQLE--SKLSLNPKNPKNVLONFVSFFNKKKEA--ETAETENTLEN	1080	
Db	1367	IKFENINETSUKNOIEIKKKKLEELN-----KNYELLAEKRETNMSISDNDKIVE	1420	
Qy	1081	TKILLKHYGKLVYNGESSPLKTLSEESIQTEDNVASLENFKVLSKLEGLKDNLNLEK	1140	
Db	1421	NNIL-----EDTDSKQNNLN--KNVEDKTGDDINCEK	1450	
Qy	1141	-----KKLSYLSGLHHLIAELKEVIKKNYTGNSPSENNTDVNNALBSYKFLPEGTDV	1195	
Db	1451	NNDAQEISYKDIKEIKISMLYGEELNRK-----NSYDEKVLNLTNEIKELK-----	1497	
Qy	1196	ATVVSSEGSDTL--EOSOPKPKASTHVGAESNTITTSQNVDDVEDDVITVPIFGSEEDYD	1254	
Db	1498	--IRNKGGEAIAELKLNKLNKEKNKSVKQNDSESSNIIITKQDGKTPEYVSNDKIQKD	1555	
Qy	1255	DLGOVVTGEAVTSPVDNITLSKIENEYEVLYL-----KPLAGVY---RSUKKQLEN	1302	
Db	1556	WKANLVILKLEKPDLDNLSLEKENFRYMSVTKENKNQVNDKIVGIYSEYFKKEKELKN	1615	
Qy	1303	NVMTFNVNVKDIILNSFRKNRFKNVLESDDLPIPKDLTSSNVVVKDPYKFLNKEKROKEL	1362	
Db	1616	DMVICVLKDIILSILF--LNDNFVNLF-----EKIDKIL	1648	
Qy	1363	SSYNYIKDSITDINFANDVLYGIKLSBKYSKSDLSISKKY-----INDKQGENEYKLPFLN	1419	
Db	1649	WKQMYIPTEI-----RILFURYFSLDKLRNVKVCNVEEYVNNERYEYSWA	1694	
Qy	1420	NIEFLYTVNDKIDLFIHLAEKVLNRYTEK-----SNVE-YKIKE-LNY-----	1462	
Db	1695	LFQTYLETAS-----NLKKMEIYVULEAKESDCENNSNFDPKPIITDILNFKSDSI	1746	
Qy	1463	-LKTIOQKLADFPKNNNFVGIADLSDYNNHLLTKF-----LSTGMVFENLAKTVLSNLD	1518	
Db	1747	RLKTI-----AQLRKEUNLFEAKNIIUNYDQIILNKYHECLRKLIKVNKARE-----LD	1797	
Qy	1519	GNLQGLMNIISOHCVKKQCPQNSGCFRHLDEREECKLLNFKQEGDKCVENPNPTCNE--	1576	
Db	1798	FN---YNVSSKFSIKKELEMCs-----DENDE-----FKYNNIKNNEEKNDOTIKDPK	1841	
Qy	1577	NN	1578	
Db	1842	HNN	1844	

Db 1813 QINYDNNLLKRLDQLFNOD-----LQKHLDTNOKLEQLKYDYIEIKERLKIETK 1867
QY 1020 DKKKTVGKYMOKK-LTLLEKQESKLANPKPHVLN-----ESVFPN 1064
Db 1868 INKQE--KYLQLOQDNHILNDFSTTTTNNNNNNNNNNNNNDNNNDTVOQFTH 1925
QY 1065 KKKEAIAETENTLENTKILLKHYGLVKYNGESSPLKTLSEESIQTEDNYASLENPKV 1124
Db 1926 SLKA-----NLNSRLLEKLSNL-----NEKQLSDE-----KNRMK 1958
QY 1125 LSKLEGLKDNLNLEKKKLSYSSGLHLIAELKEVI--KNKNY-----TGNSPSE 1173
Db 1959 ITILEDKLFKN--EKDKM-----KLOQIIDNNKNYMIQYNKLTNLDMLSE 2003
QY 1174 NNTDVNNALESYKFLPE-----GTDVATVVSSESGSDLEQ-----1209
Db 2004 ENRMILLNKEEYEQIEQLNHHDKLFISTKNNDIQIENKELOEQVDQYITINEKDKII 2063
QY 1210 ----SOPKKA--STHVGAESNTITTSQNVDDVDVIV--PIFGESEEDYDGLGVVVT 1261
Db 2064 VHLNLQIKLANQNEHMRSCDIFNVAHSQDNIKNHVMVGEDIMGDTNHDVKNKIDQGT 2123
QY 1262 GEAVTPSVIDNLSKIENEYEVL-----YKPLAGVYRSLKQLENV-----WTFN 1308
Db 2124 NOHINOGTNOHI--NOGTNOHDTCDGPNYVYKVQVQATNREDNKNKERNLSQEIYKIYINEN 2182
QY 1309 VNVKIDLSRENKRENFVNL--ESDLIPYK-----DLTSSN-----YVVKDPYKFL-- 1353
Db 2183 IDLTSELEKKDMLENYKNELEKEEYIYKLNDDMLSNCKKLEKESIMMEKYKIIMN 2242
QY 1354 -----NK--EKRDKPLSSYNYIKDSI-----DTDI--NFANDVLGYKILS 1390
Db 2243 NNIQEKDEIENLKNYKNNKLDLNNYSVVDKSIWSCFEDSNIMSPSCNDILNVFNLS 2302
QY 1391 EYK-----SDLSIKYIYNDKQENEKYLPFLNNIETLYKTVDK-----IDLVP 1436
Db 2303 KSNKKVCTNMDCNMOSISSINNYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2362
QY 1437 IHLEAKVLNVTYKSNVYKIKELNYLTIQDKLADFKNNNFVGIADLSTDYNNH-----1492
Db 2363 VN-----NLQLKNDNNIIKFNILKL-----FKLGSCLYI-----INRNLEI 2403
QY 1493 -NLLKFLSTGVFENLAKTVLSNLDGMLQGLMISQHCVKKQCPQNSGCFRHLDERE 1551
Db 2404 QMLKQILSLESIKSL--NEFINLNKEN-----EKNELIK-----INNPE 2443
QY 1552 ECKCLLNYKQEGDKVCPNPCTCEN 1577
Db 2444 EILKLNLLQDNESCIONLNNYLKKN 2469

RESULT 29

G70163

hypoetical protein BB0512 - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999

C:Accession: G70163

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: G70163

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2166 <KLE>

A:Cross-references: GB:AE001153; GB:AE000783; NID:g2688419; PIDN:AAC66876.1; PID:g268842

A:Experimental source: strain B31

Query Match

6.0%; Score 507; DB 2; Length 2166;

Best Local Similarity 20.8%; Pred. No. 1.6e-08;
Matches 355; Conservative 306; Mismatches 552; Indels 494; Gaps 88;
QY 132 VRNYLLTIKELKYPQLFDLTNHLMLTCDNIHGF-----KYLIDGVEEINELLYKLN 182
Db 9 VNLFLSVILFVYRQ--YDKRSRAL---DKIKKFPVDLTUVNLEDFIEDTKTEINDLAVDME 64
QY 183 FY-----FDLLRAKLADV-----CANDYQOI-----PFLNLKIRAN-----212
Db 65 AVORSIEIIK-KIEVQOKIKNKSNDFAEVEKKIAYHDSMLKDLDEMFKVQENIQLQ 123
QY 213 ----ELDVLUKLVFGYRKPDLNKNVKG-MEDYIKKNKKTIE-----NINEL 255
Db 124 VDGKIVDKLSKLTGFTQIDSVESNLVLEKFKDANKENLESIKIASWEKFDPTIKEL 183
QY 256 I-----ESKKTIDKNK--ATKEEEKK-----LYCAQVDLSLYNKKOLE 294
Db 184 VFKIDNLNKEISLYEKDLANIEERKNDILVKGNEKLDLEFSEFLEKVEFNIGYKSEIES 243
QY 295 AHN-----LISVLEKRIDTLKKNENIKE--LLDKINEIKNPPPPANSNGTPTNLL---341
Db 244 SFNFEYENKYKLIENTSELIMESVKNKINEKEDFILNRLNE-----ELQNKFKDILVYV 296
QY 342 -DKNKKIEBH-----EKEIKETAKTIKN-----IDSLFT 370
Db 297 DDRSKEIKDKLEDKLVLDVNEISSLSSSPKDNVYSRINSLEESMRITEMGKYEQVDVDF-355
QY 371 DPLELEYLREKKNKIDISAKVETKESTEPNCPVTPSYPLSYNDINNALNELNSFGDLI 430
Db 356 DKFRSQVQLNKL-----IYEDYEDKISQVDNNIRKV--ELSLDLNLSKMSVQS-----404
QY 431 NPFDYTK---EPSKNIYTDNERK-----KFINEIKE-KIKIEKKIESDKKSYED 476
Db 405 GAIDFKRLLEDSDNGIYLEFKGFGADIEVSESPKGDINQLKMQLESQLLDVSQIQR 464
QY 477 RKSJNDITKEKELLNIIYDSKFNNDILDTNPEKMMGRYKYKYEKLTHHTTFASYENS 536
Db 465 LIKLDNLSNFEI-----NGRFNNN--YSNLDNINAKYTALFESL--DSSSSKFENQ 515
QY 537 ----KHNEKUTK-----ALKYMEDYSURNIVVEKELKYKNLSKIEIENETLVENIK 586
Db 516 MESKYSFTDKLTAGMDEFSLMYGEKFPETLSQEAATNNYQEFQDLNKKLENEIESFNME 575
QY 587 KDEBOL---FEKIKTDENKPKDEKILEVSDIVKQVQVLLMKNKIDELK-----KTQ 635
Db 576 KTQETILKVDNTSLINIKDEIGNIVEFRDYYDEVN--IFVTQLEESKLQYSKMQGEMD 633
QY 636 LILKNVELKHNTHVPNSYKQENKQBPYIILVLKKEIDKLKVPMPKVESLINE-----EKK 690
Db 634 SNLKNI-----SQINKTNEEFUSLIQIQDK---GIELSSEFVNDLSHIQKK 679
QY 691 NIKTEGSDNS-----EPSTEGEITGOATTK---PGQAAGSALE-----GDSV 730
Db 680 AIDMGSWKDELIALNKSLLDIKVSSE-ELLSATLIKLESLEKVDNDRMEYVLLKTGDIE 738
QY 731 QAQAQEQKQAPVPVPVPEAKAQPTTPAPVNNKTE-----NVSKLDYLEKLEYFLNPTS 785
Db 739 SLVIEKYKELK--DMSYSQSDEAILGIEKFINRQETIILKDKSVFMLEDLNKKFKDDKN-N 794
QY 786 YICHK-----YILVSHSTWNELIKQYKTKTEE--ESKLSSCDPL-----DLLFNI 829
Db 795 FVISEEDCYKLKDFKIESEDILNFKSDLNEFIESKLQIVSNIKSDNQKQIDDDLDRI 854
QY 830 QNNIPVMVSMFSLNLSLSQLEWIEYKEWCVNLYKLKND---KIKNLEEAKEAK-----880
Db 855 SKDI---LNRKSNINNEVDKLSL-----WOSKLNIEITVKIENLLSSGKVDLDLI 901
QY 881 --KVSTSVKTL---SSSSMQPLSITPODK-----PEVSANDD 912
Db 902 DSEVTTIKELKFSLESLESYILEKIDEFERNOQAIYSDELLQDIMNHFNKETRELENLS 961
QY 913 TSHSTNLNLSKLFIENLSLGNK-----NIYQELIGQKSSSENFYEKILKSDST 961

Db 779 QILNSSHSLKENFISILETELKKNVRSLODEMTOQLRVLETKD--KENQTALL---EYKST 833
QY 453 INEIKKIEKKIE---SOKSYED---RSKSLNDITKE---YEKLLNEIYDSKFN 502
Db 834 IHKOEDSIKLEKLETLISQKKAEDGINKMGKDLFALSREMOAVEENCKNLOKEDKS 893
QY 503 NIDLTFEKMGRYKVKYKLEKTHHTTFASYENSKINLEKLTAKALYMEDYSLRNIVVER 562
Db 894 NVNHOKETSLKEDIAAKITEIKAIN--ENLEEMKLOCNLSKEKEH-----ISK 941
QY 563 ELKYYK-----NLJSKIENETIETVENIKKDEQOLFEEKITKDNKDPKILEVSDIV 615
Db 942 ELVEYSRFQSHDNVLAKLETKLSANNY-KDMAQENESLIKAVEESKNESIQLSN-- 998
QY 616 KVQVQVLLMNKIDELKKTQILKLVNKLHNIHVPSYQENKQEPYLLVLKKEIDK 675
Db 999 -----LQNKIDMSQE-----KENFOIERG-----SIEKNLEQLK 1028
QY 676 VFMKPVESLINEKKNIKTGQSDNSEPSTEGEITQATTKPGQAGSALGDSVQAQ 735
Db 1029 ----KTISDLEQKKEIIS--XSDSSKDEYESQI---SLLKLEKLETATTANDENVNKISE 1079
QY 736 EQQAQPPVPVPEAKAQVPTTAPVNNKTENVSKLDVLEKLYEFLNTSYICHKILVS 795
Db 1080 LTKTRE-----ELEAEL---AAYKNLKNLETKLETSEKALKEVKE----- 1117
QY 796 HSTMNKILKOYKITKEESKLSCDPLDLFNION-----NIPWYSMFD----- 841
Db 1118 ----NEHLKEEKIQLEKATETATKQOLNSRLANLESLEKEHEDLAQLKYEIQIANKE 1173
QY 842 SLNLSQOLMEIYEMVCNLYKLDKNDKIKNLLPEAKKSVTSVKTLSSSMOPLSLTP 901
Db 1174 QYNEEISQNLDEITSTQOENESIK-KKNDLEGEVAMKSTSEEQSNLKKSEIDALNL-- 1230
QY 902 QDRPEVSANDTSHSNL-----NNSLKL-----FENILSLGNKNI 938
Db 1231 QIK-ELKKNKNETNEALSLESIKSVESETVKIKELQDECNFKEVSELEDKLKAESDKNS 1289
QY 939 -YOELIGOKSENFYKILKSDST---FYNESFTNFVKSADINSN-----DESKR 988
Db 1290 KYLEL--QKSEKIKEL--DAKTELKIQLEKITNLKAKESSELSPLKTSSEERK 1345
QY 989 KLEEDINKLKTQLSFDLYNKYKYLKLERLFDKKTVGKYKMOIKLTLKLEQESKLNS 1048
Db 1346 NAEEOLEKLNKEIQIK-----NOAFEXER-----KLLNEGSSTIQEYSEKINT 1389
QY 1049 LNNPKHVLFQNSVFNKKKEAEIATENTLENTKI---LLKHVGLVLYNGESSPLK- 1103
Db 1390 LEDELIRLQN-----EHELKAKEIDNTRSELEKVSLENDELEEKONTIKSLQDEILSYKD 1445
QY 1104 --TLSEESIOT--EDNYASLENFKVLSKLEGLKLDNLNLEKKLSYLSGLHHLIAELKE 1159
Db 1446 KITRNDKLLSIERNDRKOLLESLEKQOLRAAQESKAKVEEGLKLEESSEKAELEKSKE 1505
QY 1160 VIRKNVYTGNSPENTVDVNNALLESYKFLPEGTDVATVVVSESGDTLEQSQPKKASTH 1219
Db 1506 MMKLESTIES---NETELKSSMETIRK-----SDEKLEQS--KKSAAED 1545
QY 1220 VGAESNTITTSQNVQDEVDVVIPIFGESEEDYDDLQGVVGTGEAVTPSVIDNLSKIEN 1279
Db 1546 I-----KNLOHEKSD--LISRINSEKIDIEELKSLKRIEAKSGSELETVTKQELNN 1593
QY 1280 EYEVLYL-----KPLAGVYRSLK-KOLENNVMTFNVNVDILNSRKNREKNVNL 1329
Db 1594 AQEKIRINAEENTVLSKLEIDIEBRLKDKQAE---TKSNQEEKELTSLRLEKE----- 1644
QY 1330 ESDLIPYKDLTSSNYVVKPYKFLNKEKDKFLSSNYNFKDSIDTDFINFA-----NDVLGY 1385
Db 1645 -----QELDSIQQAOK-----SEERAEVRKQVEKSQOLDEKAMLETKYNDLVN- 1691
QY 1386 KYLSEKYSDDLISIKKYINDKOGNEKYLPLFNLTIE---TLYKTVN-----DKIDL 1436

Db 1692 ---KEQAWKRDEPTVKKTTDSQRQEIEKLAKELDNLKAENSKLKEANEDRSIEDDLMLLV 1748
QY 1437 IHLEAKVLNVTYEKSNVEVKI 1457
Db 1749 TDLDERNAKYRSKLKDLGVEI 1769
RESULT 33
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T28317
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; MUID:99102612; PMID:9847359
A:Accession: T28317
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFO>
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97677.1; PID:g4049717
C:Genetics:
A:Note: MSV156
Query Match 5.7%; Score 483; DB 2; Length 1127;
Best Local Similarity 19.7%; Pred. No. 4.2e-08;
Matches 287; Conservative 247; Mismatches 430; Indels 490; Gaps 65;
QY 132 VRNYLTIRELKY-PQLFDLTNHLMLTCDNHGFKYLDGYEINELLYKLNFEYDILLRA 190
Db 54 IKNYNIVEIKFYMHQFQKINDYNIIL-----QYLLIYNEINKEICKENKF----- 99
QY 191 KLDVCANDYQIPFNLKIRANELDVL---KKLVFGYRKPLDNKIDNKGMDYIKKKK 246
Db 100 ----PCKNPLYNTYKKKLYIDLYEKKDKELVINIQ-----KNVDKIND-IKNNV 149
QY 247 KTIENINELIEESKKT-----IDKNKNATKEEEKKLYQAQYDLSIYNKQLEEAHNLISVL 302
Db 150 NNTHSDNETIITGKETLIDILNLKLLVSDSEKQLEIQYKNN--NKEIE----- 198
QY 303 EKIDILKKNENIKELLDKINEIKNPPANGNTPTNLLDNKKIEHEKEIEIAKTIK 362
Db 199 FRNIDNVQREINKQ--DELN-----LLDESKK--EFIKQOEELNKTID 239
QY 363 FNIDSFLF--TDPLEYLEYLRKKNID-ISAKVETESTEPNEYPNGV--TYPLSYNDIN 417
Db 240 KQOEELIKKLNDEINFNIDKQKLLDQINSKINTL-----NENIKGVMLYETKKNIS 294
QY 418 NALNELNSFGDILNPPFDYKPSKNIYTDNERKFFINEIKEIKIEKKIESDKKSYEDR 477
Db 295 NLQNEILNKDSTIKSLD-----EKOQLLDEL----- 320
QY 478 SKSLNDITTEYKELLNEIYDSKFNNDILNFEKMGKRYSKYKVELTHHTNTFASYENSK 537
Db 321 DKNNILNITSYLNK-----SNTKIINIQQLL-----ESSLTDFFNNAN 356
QY 538 HNLEKLTALKYMEDYSLRNIVVVEKELKYKLNLSKTIENEI---ETLVENIKKDEEOLFE 594
Db 357 ININEL-----KSKIKLFDNDIQKLNNDITEQNNKTIOTFFENNSTRIFK 399
QY 595 KITTKDENKPKDEKILEVSDIVKVOQVQLMKNIDELKKTQILKLVNKLHNIHVPSYK 654
Db 400 EKLDTE-----YKKIDIDIKNNN--LOKLE-----ESYK 425
QY 655 QENKQEPYLLVLKKEIDKLKVFMPKVESLINEEKKNIKTGQSDNSEPSTEGEITQAT 714
Db 426 KIDEQTEYI-----KNKINKEYNDI----- 445
QY 715 TKPQQAGSALGDSVQAQAEQKQAOQPPVPVPEAKAQVTPPAPVNNKTENVSKLDY 774
Db 446 -----TELKNNLOKLEENKK-----IDEGTEY-----Y 470


```

OY 996 -----LKSDFTYNEEFT-----NFVKSADIDNSL-----NDESKR 987
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1178 MDVELQDARTKLDNSRVKKNHNSITIQOHDIMKLNJNLRESNITLRLNELENNNNKK 1237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 968 KLEEDIKRLAKTLL-----QSFIDLNR-YKIKLERLPDKKTKYGRWQIKKLTLL 1038
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1238 KeloSELDKLNQNVAPIESELTALKYSMQEKEQELT-----AKEEVHRMKKRSQDILEK 1292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1039 KEQLES-----KLNS-LNNPKHYLN-----FSVFPRK-KKEAE-----IAE 1073
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1293 HEQLSSDYEKLESEIENIKLEELENKEROGAEBEKEFRRLRQAOERLTKLSLQSDSLTE 1352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1074 TENTLENTKILKLHKGLVYKYYNGFSPLTKLSESEISTEDNYASLENFKVL---SKLEG 1130
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1353 QVNSLRDKNVLEN-----SLSENNAPIEE-----LQNAKVAQGNQOLEA 1392
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1131 --KLKDN-----LNLERKKLSYLS-SGLHLHIAELKEVIK-----KKNYGNSSPS 1172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1393 IRLIQEDAEKASRELQAKLESTSTYESTINGLNEIITLKEETIKQROQOQLQATSAN 1452
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1173 ENNDVNNALSYKKFLPEGLDQVATVSESSDILE--QSQPKKPASTHYGAESNTITTT 1229
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1453 EON-DLSNIVESMKKSFE--DKIKFIKEKTOEVNEKILAEQERLNQPSNNIMEIKKKW 1509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1230 SONVDEDDVYIYIFESSEDYDDLGQVYTGAVTSPSVIDNLSKLENEYEVYLYKPL 1289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1510 ESEHEQEVSOKT-----KEAEALKKRIRLPTEKIN-KIERKKEELEKEFEKEVEERI 1563
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1290 AGVYRS-----LKKOLENNYTFENVNVKILNSRFNRENFKNVYLSDLIPYKDLTSSN 1343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1564 KSMQSGEIDVLRKLOPAKVO---EKOKELENEYKK-----LQEL--KDVPHSS 1610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1344 YVVDYPRFLNKKERDKRLSSYNIKDSIDPDINDINANVLGYKILSKYSQSDSTIKKY 1403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1611 HISDD-----EKDK-----LRAEIESR-----LRREFNNELQAIKK 1642
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1404 INDKGENEKYLPFLNNITETLYKTVNDKIDLFVILHLEAKVLYNTEKSNVEYKIKELNYL 1463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1643 SFD-EGKQ-----QAMMKTP-----LLEKR-----L 1662
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1464 KTIDQKLADLFKKNNFVGIADLSTIDYNNINNLTKLSTGVAFENLAKTVLSNLDGNIQG 1523
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1663 AKMESQLETSKQSA-----SEPKSVNNVONPLGLPRKIEENSFPFPLSG----- 1711
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1524 MLNISQHCYVKQCPQNSGCFRHLDERECKCLILYKQEGKCVENPNP 1572
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1712 -----EKLLKNSKSSSGGFNPFTSPSPKHLONDNDKRESLANKTDP 1754
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 37
B44997
merozoite surface antigen 1 - Plasmodium chabaudi chabaudi (strain CB)
C:Species: Plasmodium chabaudi chabaudi
C:Date: 30-Sep-1993 #sequence: revision 30-Sep-1993 #text: change 11-Jan-2000
C:Accession: B44997
R:Lew, A.M.; Beck, D.J.
Mol. Biochem. Parasitol. 42, 153-154, 1990
A:Title: The epitope of a protective monoclonal antibody occurs in a region of microheterogeneity
A:Reference number: A44997; MUID: 91042831; PMID: 1700297
A:Accession: B44997
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-476 <LEW>
A:Cross-references: GB:M43255
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 5.5%; Score 465; DB 2; Length 476;
Best Local Similarity 27.0%; Pred. No. 5,8e-08;
Matches 131; Conservative 78; Mismatches 145; Indels 132; Gaps 12;

672 DKLAVMKRVESLINEEKNKIKTGG-QSDNSEPSTEGEITGQATTKPGQA----- 721

```

```

Db 1 ERLAKFIKIDNMIEKQKMEQEHVATGSEQASASAGTSSSTETTSQTPAPVAPAP 60
Oy 722 -----GSALEBDSV---QAQAEQEKQAP-----PVVPVPEA 751
Db 61 AEKAEKGETESTETPTPAASKPAEGAASTGATPTPEQEAAPTEQEAQAPAPETPAEVPAPT 120
Oy 752 KAQVPTPAAPVNNNTENVSKLDYLEKLEYEFLNTGYICHKYLIVSHSTMNEKILKQYKITK 811
Db 121 PAAPATPAAPAPAPAKPVWTKLYLEKLEKFLAFPSYCHRYVLONSTINKDALSKEYALTP 180
Oy 812 EEES--KLSSCDPLDLF-----827
Db 181 EEDKIRTLKRCSELDVLLYKAGVIDPEVPAPTAPVAPETAPETAPETPAQEAPOQ 240
Oy 828 ---NIQNNIPVYMSFSDLSNLSLSOLFMEIYEKEMVCLYLRD-NDKIRNLEBAKVS 883
Db 241 PESAIQNNMPMYSLYENVVGLQNIYTELXEKEMMYIYNLKKPNAPVAKALLAQAEPA 300
Oy 884 TSVKT-LSSSMOPLSLTPQDKP-----EYSA 909
Db 301 TETTTTTPASASTETPTPKAPTATPTSEYVTOGCTTPAAPKAQEGASSAPAPAPAKAPA 360
Oy 910 NDDTSHSTINLNSLKLEPNILSLGKNKNIYQELIGQSSSENYEKILKSDTFYNESEFTN 969
Db 361 QTVGQSTNVGSGYV-----RAESEDMEFVDFEVDNFKSYLQYDVG-NTQFID 411
Oy 970 FVKSADDINLNDSESKKKLEEDINKTKTLQLSFPLNKKYKTLKRLRFPKKTGYK 1029
Db 412 FIKSKELMLNLTPE-KVNLTYLDIAHLKELSEHYNNRYKKLEKLELYQHBOLEAAN 470
Oy 1030 MQIKKL 1035
Db 471 QKVKET 476

RESULT 38
A:44997
merozoite surface antigen 1 Plasmodium chabaudi adami (strain DK)
C:Species: Plasmodium chabaudi adami
C:Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 11-Jan-2000
C:Accession: A44997
R:Lew, A.M.; Beck, D.J.
Mol. Biochem. Parasitol. 42, 153-154, 1990
A:Title: The epitope of a protective monoclonal antibody occurs in a region of microh
A:Reference number: A44997; MUID:91042831; PMID:1700297
A:Accession: A44997
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: DNA
A:Residues: 1-482 <LEW>
A:Cross-references: GB:M34047
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 5/58; Score 460; DB 2; Length 482;
Best Local Similarity 26/28; Pred. No. 8,4e-06;
Matches 129; Conservative 78; Mismatches 147; Indels 138; Gaps 10;

Oy 672 DKLKVFMPKVSLLNEEKKNIKTE---GQSDNSEPSTEG-----707
Db 1 DRLAKFIKIDMTEKEQKMEQEPVATGESEQVTPSSGAGTSTQTAQTTQTPPAAPAPV 60
Oy 708 ---ETTGQATTKPGQQAAGSALGEDSVQAQ-----AQEQKQAPPVVPVPEAKAQVPTP 758
Db 61 KETTEITPTKATQEPQADDTTTPEGTTPTQEGAAAPTAEPQAPATETPAEVAVPAPAPAP 120
Oy 759 PAPVNNKTENVSKLDYLEKLEYEFLNTGYICHKYLIVSHSTMNEKILKQYKITKEES--K 816
Db 121 AAPAPAPAPVWTKLYLEKLEKFLAFPSYCHRYVLONSTINKDALSKEYALTPPEDKIRT 180
Oy 817 LSSCDPLDLF-----827
Db 181 LKRCSELDVLLYKAGVIDPEVPAPTAPVAPETAPETAPETPAQEAQEAPOQ 240

```


A:Gene: GDB:CEPHE
 A:Cross-references: GDB:361164; OMIM:117143
 A:Map position: 4q24-4q25
 C:Superfamily: centromere protein E: kinesin motor domain homology
 C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
 F:7-35/Domain: kinesin motor domain homology <KNOT>
 F:86-93/Region: nucleotide-binding motif A (P-loop)
 F:486-2183/Domain: coiled coil #status predicted <COI>
 F:92/Binding site: ATP (Lys) #status predicted

Query Match 5.38; Score 443.5; DB 1; Length 2663;
 Best Local Similarity 18.48; Pred. No. 1.9e-06;
 Matches 345; Conservative 349; Mismatches 617; Indels 559; Gaps 82;

```

QY 126 ADKRRVRYNLTITKEL---YPLFDLTNNHMLT-----LCQNIHGFXY 166
D 126 ADKRRVRYNLTITKEL---YPLFDLTNNHMLT-----LCQNIHGFXY 166
QY 415 AKRRRTWGLGKINRKNNSYADQFNPITTKIKLSINLREIDESCSDVFSN 474
D 415 AKRRRTWGLGKINRKNNSYADQFNPITTKIKLSINLREIDESCSDVFSN 474
QY 167 LIDYEELI---NELLYKLNFFDL--LRKLVDCANDYCOI---PPLKIR-ANE 213
D 167 LIDYEELI---NELLYKLNFFDL--LRKLVDCANDYCOI---PPLKIR-ANE 213
QY 475 TLDTLSIEWNPATKLLNQNIESELNLRADYDNLVL--DYQLRTEKEEMELTKRND 533
D 475 TLDTLSIEWNPATKLLNQNIESELNLRADYDNLVL--DYQLRTEKEEMELTKRND 533
QY 214 LDV-----LKKLVFG---YKRPIDN-----IKDVGK 237
D 214 LDV-----LKKLVFG---YKRPIDN-----IKDVGK 237
QY 534 LDEPFALEKRTKKDOEMQLIHEISNLKLVNHRVYMODLENEISKVELREKEDIK 593
D 534 LDEPFALEKRTKKDOEMQLIHEISNLKLVNHRVYMODLENEISKVELREKEDIK 593
QY 238 MEDYIKRKKKTIEININ-----ELIEESK-----TID-KNKVATKEEEKKL 278
D 238 MEDYIKRKKKTIEININ-----ELIEESK-----TID-KNKVATKEEEKKL 278
QY 594 LQEVDSQK--LENIKMDLSYSLESDIPKQMTLPDAETVALDARRESAFLRSEML 651
D 594 LQEVDSQK--LENIKMDLSYSLESDIPKQMTLPDAETVALDARRESAFLRSEML 651
QY 279 YQAOYDLSYINKOLEEHNILSVLEKRIIDPLKN--ENIKELDKTIEIKNPPRANGNT 336
D 279 YQAOYDLSYINKOLEEHNILSVLEKRIIDPLKN--ENIKELDKTIEIKNPPRANGNT 336
QY 652 KKKKKELATYKQME---NDIQYQSQLEAKKKMQVDLEKELQSAFNEIKRLSLIDGKV 708
D 652 KKKKKELATYKQME---NDIQYQSQLEAKKKMQVDLEKELQSAFNEIKRLSLIDGKV 708
QY 337 PNTLL-----DKNKRIEHE-----KIKELAKTKI--- 362
D 337 PNTLL-----DKNKRIEHE-----KIKELAKTKI--- 362
QY 709 PKDLCLMLEEGKITDQKELNKEVEENALREVILLSELKSLPSEVERLRKTIQKSE 768
D 709 PKDLCLMLEEGKITDQKELNKEVEENALREVILLSELKSLPSEVERLRKTIQKSE 768
QY 363 -----FNIDSLTDPLELEY---YLREKNKNIDISAKVETKESTEPNEVPGVYPLS 412
D 363 -----FNIDSLTDPLELEY---YLREKNKNIDISAKVETKESTEPNEVPGVYPLS 412
QY 769 ELHITSEKDLSEVYVHKESRVQGLLEIGTKTDLATQSNYKSLDQEFQNKTLMD 828
D 769 ELHITSEKDLSEVYVHKESRVQGLLEIGTKTDLATQSNYKSLDQEFQNKTLMD 828
QY 413 YN-----DINNALNL-----NSPGDLINPDY--TKESKNIYDNEKKF 452
D 413 YN-----DINNALNL-----NSPGDLINPDY--TKESKNIYDNEKKF 452
QY 829 FEQKYKVALEENEMNOEIVNLSKQAKFDSGLALTELSTYKQELQOETREVOEYLN 888
D 829 FEQKYKVALEENEMNOEIVNLSKQAKFDSGLALTELSTYKQELQOETREVOEYLN 888
QY 453 INEIKKIK-----IEKKK-----IESDKSYEDRSKSLNDITKEYEKL 491
D 453 INEIKKIK-----IEKKK-----IESDKSYEDRSKSLNDITKEYEKL 491
QY 889 MEQLKQLENRDSPLOQVEKEKTLITTEKLOOTLEEVKTLTQEKDQLQDOESLDIERDQL 948
D 889 MEQLKQLENRDSPLOQVEKEKTLITTEKLOOTLEEVKTLTQEKDQLQDOESLDIERDQL 948
QY 492 LNEIYDSKFNNNIDLTNEFKMMGRYSYKVEKLTNN--NTFAS-----YEN 535
D 492 LNEIYDSKFNNNIDLTNEFKMMGRYSYKVEKLTNN--NTFAS-----YEN 535
QY 949 KSDIHDT--VNMNIDTQ-----EQLRNALSLEKQHOETINTLSKSISEEVSRNLHMEEN 1000
D 949 KSDIHDT--VNMNIDTQ-----EQLRNALSLEKQHOETINTLSKSISEEVSRNLHMEEN 1000
QY 536 ---SKNLEKLTAKLAKMEDYSLRNT-----VEKELKYTNLISKINLETETL 581
D 536 ---SKNLEKLTAKLAKMEDYSLRNT-----VEKELKYTNLISKINLETETL 581
QY 1001 TGEFKPEFOQKMWGIDKQDLKAKNTQTLTADYKDNELIISQQRKF--SLIQE--KNEIQOM 1058
D 1001 TGEFKPEFOQKMWGIDKQDLKAKNTQTLTADYKDNELIISQQRKF--SLIQE--KNEIQOM 1058
QY 582 VENIKDEEOLFEEKIKTDEN-----KPDKEILEVSDIVVOYQVLLNMRKIDELKTKQTL 636
D 582 VENIKDEEOLFEEKIKTDEN-----KPDKEILEVSDIVVOYQVLLNMRKIDELKTKQTL 636
QY 1059 LESTIAKEBOL---KTDLKEIKETIENQEBELRLLSDELKKQ--OEIYAQOEKNHAIKEGE 1114
D 1059 LESTIAKEBOL---KTDLKEIKETIENQEBELRLLSDELKKQ--OEIYAQOEKNHAIKEGE 1114
QY 637 ILKN---VELKNIHIVPNSYKQENKQEPYLLIYLKREIKLAKFMPVSLINEEK--- 689
D 637 ILKN---VELKNIHIVPNSYKQENKQEPYLLIYLKREIKLAKFMPVSLINEEK--- 689
QY 1115 LSKRCPLALAEVEKLKRSQQLQKQO---LNVQSEMESEMKKIMEIKELKLNKKE 1171
D 1115 LSKRCPLALAEVEKLKRSQQLQKQO---LNVQSEMESEMKKIMEIKELKLNKKE 1171
QY 690 ---KNIKTE---GQSDNSEPSTEGELTGOATTKPGQOAGSALEGD-----SVQAOAQ 735
D 690 ---KNIKTE---GQSDNSEPSTEGELTGOATTKPGQOAGSALEGD-----SVQAOAQ 735
QY 1172 LTLEHMETERLELAOKLNENVEEYKSTIKERKVLKELQKSPETRDHLRGYIRIETATGL 1221
D 1172 LTLEHMETERLELAOKLNENVEEYKSTIKERKVLKELQKSPETRDHLRGYIRIETATGL 1221
QY 736 EQQAQAPVPVPPVPEAKAQVPTPPAPVNNKTEVNSKIDYLEKLEYFLNTSYICHKITYLS 795
D 736 EQQAQAPVPVPPVPEAKAQVPTPPAPVNNKTEVNSKIDYLEKLEYFLNTSYICHKITYLS 795
QY 1232 QTKEELKIAIHILKHEOETIDELRRSVSEKTAQIINTQDLEK-----S 1274
D 1232 QTKEELKIAIHILKHEOETIDELRRSVSEKTAQIINTQDLEK-----S 1274

```

```

QY 796 HSTWNEKILKQVITKEESKLSQDPLDLLFNIONIPVYMSFSDIANS----- 846
D 796 HSTWNEKILKQVITKEESKLSQDPLDLLFNIONIPVYMSFSDIANS----- 846
QY 1275 HTKIQDEI-----PYLHEQELLPVKKVSVSETQETMNELELLTQSTKDTTLARIEMER 1330
D 1275 HTKIQDEI-----PYLHEQELLPVKKVSVSETQETMNELELLTQSTKDTTLARIEMER 1330
QY 847 --ISQFMETIEYKEMVCNLTLLKDN-----DKTKNLLEAKKAVSYSTYLTSS 892
D 847 --ISQFMETIEYKEMVCNLTLLKDN-----DKTKNLLEAKKAVSYSTYLTSS 892
QY 1331 LRINEKFOE--SOEELISLTKERDNLKTIREALEVKHDKOLEHIRE-----TLAKIOESO 1383
D 1331 LRINEKFOE--SOEELISLTKERDNLKTIREALEVKHDKOLEHIRE-----TLAKIOESO 1383
QY 893 SMDPLSTPODKPEVANSNDQTSHTNLSNLSKLE-----NILSLGKNKIYQELI 943
D 893 SMDPLSTPODKPEVANSNDQTSHTNLSNLSKLE-----NILSLGKNKIYQELI 943
QY 1384 SKQDQSLNKKK-----DNETTKIVSEMDQKPDQSLALREIEMLSKRL----- 1430
D 1384 SKQDQSLNKKK-----DNETTKIVSEMDQKPDQSLALREIEMLSKRL----- 1430
QY 944 GOKSSENFYEKILKSDPTFFN-----ESPTNFYKSKADDINSUNDESKR-----KK 989
D 944 GOKSSENFYEKILKSDPTFFN-----ESPTNFYKSKADDINSUNDESKR-----KK 989
QY 1431 --QESHDEMKSVAKERDQLRQEVLOESDQKLENKEIVAKHLETEEBELKVAHCCKE 1488
D 1431 --QESHDEMKSVAKERDQLRQEVLOESDQKLENKEIVAKHLETEEBELKVAHCCKE 1488
QY 990 LEEDINKLK-----KTLQSLFDLYN--KTKLKERLPDK-----KTVGKYKMQIK 1033
D 990 LEEDINKLK-----KTLQSLFDLYN--KTKLKERLPDK-----KTVGKYKMQIK 1033
QY 1489 QEEETINELRVNLSKEKETESTIQOLEAINDKQNKIOEYIEKEEOLNIQOISEVQENVN 1548
D 1489 QEEETINELRVNLSKEKETESTIQOLEAINDKQNKIOEYIEKEEOLNIQOISEVQENVN 1548
QY 1034 KLTLLKE-----QLESKLSLNPNKPHVLQNFVSFFNKKK-----EAEIATEN 1076
D 1034 KLTLLKE-----QLESKLSLNPNKPHVLQNFVSFFNKKK-----EAEIATEN 1076
QY 1549 ELKQFKHKKAKDSALQISKMLELTNRLOESQEEIQIMAKEKEEKRVQOALQIERDQ 1608
D 1549 ELKQFKHKKAKDSALQISKMLELTNRLOESQEEIQIMAKEKEEKRVQOALQIERDQ 1608
QY 1077 TLENTKILKHYGLVYKYNGESSPLKTLSEESI--QTEDVYASLENFVLSKLGKLDN 1135
D 1077 TLENTKILKHYGLVYKYNGESSPLKTLSEESI--QTEDVYASLENFVLSKLGKLDN 1135
QY 1609 LKEETKEI-----VAKKESQOEKEYQFLKMTAVNEQEKCEIEHLKEQFQKULLEN 1662
D 1609 LKEETKEI-----VAKKESQOEKEYQFLKMTAVNEQEKCEIEHLKEQFQKULLEN 1662
QY 1136 LNLKKSLSYSSGLHLLLELKEVITKNKNYTGSPENNVDVNNALSEYKFLPEGTQV 1195
D 1136 LNLKKSLSYSSGLHLLLELKEVITKNKNYTGSPENNVDVNNALSEYKFLPEGTQV 1195
QY 1663 IETENIRLTOI--LHNLLEEMRSVTKERD-----DLASVEETLK--VEEDQL 1705
D 1663 IETENIRLTOI--LHNLLEEMRSVTKERD-----DLASVEETLK--VEEDQL 1705
QY 1196 ATVYSESGSDLEPOSQPKPASTH-----VGAESNTIITSQNVDEVDV- 1240
D 1196 ATVYSESGSDLEPOSQPKPASTH-----VGAESNTIITSQNVDEVDV- 1240
QY 1706 KENLRETITTDLEQEBELKIVHMLKHEOETIDKLRGIVEKEKTEINSMQKOLEHSDAL 1765
D 1706 KENLRETITTDLEQEBELKIVHMLKHEOETIDKLRGIVEKEKTEINSMQKOLEHSDAL 1765
QY 1241 -----ITVPIGSEEDYDLDGOVYVGEAVTPS-----VIDNIIISKIE- 1278
D 1241 -----ITVPIGSEEDYDLDGOVYVGEAVTPS-----VIDNIIISKIE- 1278
QY 1766 KAODLKIOEBELRIAHMLKQOEETIDKLRGIVSEKTDKLSNMOKDLENSNAKLOEKIOEL 1825
D 1766 KAODLKIOEBELRIAHMLKQOEETIDKLRGIVSEKTDKLSNMOKDLENSNAKLOEKIOEL 1825
QY 1279 --NEYEVLTK-----PLAGVYRSLLKQLENNVMTF-----NVNVDILN----- 1316
D 1279 --NEYEVLTK-----PLAGVYRSLLKQLENNVMTF-----NVNVDILN----- 1316
QY 1826 KANHQILTLKQDVNETQKXVSEMEQLKQIKQDSLTLKLELENLMLAOLHLENLEEMK 1885
D 1826 KANHQILTLKQDVNETQKXVSEMEQLKQIKQDSLTLKLELENLMLAOLHLENLEEMK 1885
QY 1317 SRFNKRENFKNVLESDLIPKRDLTSSNVVVPYKFLNKKRQKFLSSYNYINDSIDTDI 1376
D 1317 SRFNKRENFKNVLESDLIPKRDLTSSNVVVPYKFLNKKRQKFLSSYNYINDSIDTDI 1376
QY 1886 SVMKERNLNRVEET-----LKLERDQKES--LQETKARDL 1920
D 1886 SVMKERNLNRVEET-----LKLERDQKES--LQETKARDL 1920
QY 1377 NFANDVLGYTKILSEKYSKSDLSIKKTIINDKQENKYLEPFLNNIETLYTVNDKIDLFV 1436
D 1377 NFANDVLGYTKILSEKYSKSDLSIKKTIINDKQENKYLEPFLNNIETLYTVNDKIDLFV 1436
QY 1921 EIQOE--LKTARMLSKHEKEFVNDKLRKISERTIO-----ISDIQKDLKSDQELQKI 1972
D 1921 EIQOE--LKTARMLSKHEKEFVNDKLRKISERTIO-----ISDIQKDLKSDQELQKI 1972
QY 1437 IHLKAVLANTYKESNVEYIKELNYLKTIODKLADKKN-----NNFVGIAIDSTQY 1489
D 1437 IHLKAVLANTYKESNVEYIKELNYLKTIODKLADKKN-----NNFVGIAIDSTQY 1489
QY 1973 QELQKKEQLQLRKVEDVNMHSHKKTINEMQLLKQ--FEPYTLCKCEMDNQLTKKLHESL 2029
D 1973 QELQKKEQLQLRKVEDVNMHSHKKTINEMQLLKQ--FEPYTLCKCEMDNQLTKKLHESL 2029
QY 1490 NNNNLITFK-----FLSTGVFENIAKTVLNS-----LIDGNLQGM 1524
D 1490 NNNNLITFK-----FLSTGVFENIAKTVLNS-----LIDGNLQGM 1524
QY 2030 EETRIVAKEDDELRIKESLKMENDQFIAT--LRMLIARORQNHQVYKPERKLLSDGOHL 2087
D 2030 EETRIVAKEDDELRIKESLKMENDQFIAT--LRMLIARORQNHQVYKPERKLLSDGOHL 2087
QY 2088 MESLRKCSRIKELKKR-----YSMDHGYCLANRLSDLEKETEFRHMKKLYVLSTYV 2143
D 2088 MESLRKCSRIKELKKR-----YSMDHGYCLANRLSDLEKETEFRHMKKLYVLSTYV 2143
QY 1560 --KQEGDKCV 1567
D 1560 --KQEGDKCV 1567
QY 2144 KIKEQHECI 2153
D 2144 KIKEQHECI 2153

```

RESULT 43
 B71612
 hypothetical protein PF0555C - malaria parasite (Plasmodium falciparum)

OY 1574 CENNNGCCADACCTEEDSSNGKRTIC-----ECTKPSYPLFDGIFCCSSNFL 1623
DB 1857 DKNITINIKDDKNITNMKKNNKMYLTILMYSQECSS--FYYSIFNTLI-NDYNFL 1909

RESULT 44

A64224
hypothetical protein MG218 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 07-Dec-1999

C:Accession: A64224

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;

M.; Fumrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.

, C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MUID:96026346; PMID:7569933

A:Accession: A64224

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1805 <TIGR>

A:Cross-references: GB:U39699; GB:I43967; NID:g1045903; PID:g1045905; TIGR:MG218

A:Experimental source: strain G-37

C:Genetic code: SGC3

C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Query Match 5.2%; Score 440; DB 1; Length 1805;

Best Local Similarity 19.7%; Pred. No. 1.5e-06;

Matches 334; Conservative 301; Mismatches 566; Indels 492; Gaps 79;

OY 133 RNYLTLT-----KELKYPQLFDLNMHMLTCDNIH----GEKYL 167
DB 190 KNYLLNTIDLYNELDLENOKRLLSIEYENTYRELVSADNELQNYENIDQNOIOEKHQ 249
OY 168 IDGV-EEINELLYLNF-----YFDLLRAKLNDVCANDXCQIPNLIKIANEL-DYLK 218
DB 250 YQYRDELSQLERKIQLTKOELVDKESALRYKIDA-----DFYINLAELADYAK 301
OY 219 KLVGYRRPLDNI-KDNVGMEDYIKRKNKTIENINELIE-----ESKRTIDKKNATKE 272
DB 302 QLSF-----QDGIYKQNAQNHEDLVALNKKEDRLNTOKEFMYLRQSALIDIK----- 351
OY 273 BEKKKLYOQYDLSYINKOLEAHNLISYLEKRIIDLTKNE-NIKELLDKINETKNPPA 331
DB 352 -----LQENELFAKHLEHQNEFE--OKOSDSLKLETEYKALQHKINERKN----- 397
OY 332 NSGNTPTMTLDKKNKEHEHEKEIKETIKFNIDSLFTDPLELELYLREKNNKIDISAK 391
DB 398 ESATKSEELLQERELEFEKRE-----IDTLTLO-ASLEYEHORESSQLLDKQ 445
OY 392 VETKESTEPNEYRPGVYTPLSYDINNALNLSFGDLINFDYTKPSKNITYDNERK- 450
DB 446 NEVQOHQNELEY-----AKKELDKERNLL-----DQCKV 475
OY 451 --KFINEIKERIKIEKKKIES--DKSYEDRSKSLNDITKEYEKLINLETYDSKFNNNID 505
DB 476 DSEALIFOLKEKVAQERKELELYLVKKQKODKE--NELL-FFEKQKQ--HQADFEMELE 531
OY 506 LTNEKMMGRYSYKVKLTHHNTFASYSKHNLEKLTAKLXMDYSLNIVYERELK 565
DB 532 AKQOELFEAK--HALER-----SFILDEKDLN--TKAQOLANES-----QLK 573
OY 566 YKNNLISKINETETLENIKKDEQOLFEEKITKDN-----KPEKILEVSD-- 613
DB 574 TDKSKSADPELMQNEVENLQOEKQKLFQERTYPERNAAVLSNLOQKRELLDQOKELD 633
OY 614 -----TVQVQVLLMKNIDELKLTQILLKNEVLEKHNINHPVSY 653
DB 634 QLTGSPQERLINOHEKELVAVSEKQKELIGKLIQDFOSQTLASAKNLAEEREMAIKFE 693
OY 654 KQKKQPPYLLIVLAKKIDKLKVMKPVESL--INEKKNIKIEGQ-----SDNSEPS 704
DB 1714 -----LTKKVNQHN 1722

DB 694 KEIATEKQILL-----NDVNAEVIQADLQINSLQNERSELQAKORLADFHNDISLKL 749
OY 705 TEGETGQATTPPCQAGSALEGDSVOAQA-----OEOKAQA--QPPVVPVPEA 751
DB 750 NEVELSLQKRLQELQTLLEANQKQSYQNAQAFEBEGLDKLNKQAFNLRRKQOMEDAI 809
OY 752 KAQVTPPAPVNNKTEVNSKIDYLEKYLEFNTSYICH-----KIYL 793
DB 810 KORLSD-----KHQALNMQOAEILRKTHF-LNNAFLNHNDQOKSLQDOLATVKEFKQKID 863
OY 794 VSHSTNMEK-----LLKQYTKIEEBSLSSD-PLDLFNQIN 831
DB 864 LERSALLEKQREPEENAVGFKRHMSNKTSQLQKTEYELTKQESQTOKELEKIAF--S 920
OY 832 NIPVMSFDSLNSLSQLEMEIYEKEMVNCYLKDKNDKIKNLEBAKVSYSVKTLS 891
DB 921 DLQGDYQVFELOKQ--QEFROIAKQ-----RELDKL-----AEK-NNQVKLELD 962
OY 892 SSMQPLSLTPQDKPEVSAN-DQTSHTNL-----NNSL-----KLEFENILSLG--K 934
DB 963 NRPQALQNOKQDVOAOLELEREHOHLNEQTAENQANESLQKQREQLTKKIQAFHYELK 1032
OY 935 NKNIYQELIGKSSSENFYEKILDKSD-----TFYNESPTNFVKGADQINSLAND----- 983
DB 1023 KRNOFLALKGRFLAKEDQDQKQDELINWRKQFEKETTDEDAKKKELELEKIRSL 1082
OY 984 -----ESKRRKLEED--INKLKTQLQSLFDLYNKKYKLERLEDKKRTVGY--RMQI 1032
DB 1083 QSNVELEKREKRLATDFTNLKKVOHNOINRDQINS--QIRQFLERKKNRQFSNEANA 1139
OY 1033 KKLTLLEKQLESKLSLNPNRHNVLQNSVFPNKKKEAETATENT--LENTKILLKHY- 1088
DB 1140 KRAFLIK-RLRSFASNKLQKLEALQIKLEFDKDEQOKKELQOATLQLEQFEKKNQFD 1198
OY 1089 --KGLY-----KYNGESSPLKTSSESI-----QTEENVASLE 1120
DB 1199 TEKORLVAIKTQCEKLSDEKAKALNOKLVELKNLSQTYLANKNAEYSQOQLOQKYNLL 1258
OY 1121 NFK-----VLSLEGLKLDNLNLEKKKLSYLSGSLHLLAELKEVIAK 1164
DB 1259 DLKENLEKTRQDLQKKHRSIFARLT-KFANDLREKQKQLAKQRIYVDKNNRLKENERNL 1317
OY 1165 NYTGNSESENNTDVNNALLESKFLPBGIDVAIVYVESGSTLQOSQPKKASHVGAES 1224
DB 1318 HFLSNETERKRAVLEDDISYEKQKQATD-AIASH-----KEVKKKEGLQKLVEL 1370
OY 1225 NTFITSONVDEVDVLIIVPIFGSESEYD-----DLQGVNTEAVTPSVIDNLSKI 1277
DB 1371 ETRKTKLND-----FAKFSQREFEFENQRLKLELQKLTQOTQNSNNFKTAQOI 1422
OY 1278 ENEX-----EVLVLRP-----LAGVYSLKQLENNVMTFNVNVD-----ILNSR 1318
DB 1423 ENSYKRGMEELNFOCKEEDKRSRLYEYFRKMRDEIERKESQVLYLKETORKANLLEAQ 1482
OY 1319 FNKRENKRNVL--ESDLIPK-----DLTSSNVVADPYFLKKEK-----RQKFL 1362
DB 1483 ANKNLIEKNITDEKELKAKKDVODIDSTNQKRELNLNBNKLQLOSLIEREAI 1542
OY 1363 SSYNYIKDS-----IDTINPAND-----VLGYIKLSKEYKSDL-----DSIK 1402
DB 1343 NS-----KDSLKNKIEITIKROLHDKEMRVLRVLVDMKLAIEKQYTEINRLRTQTFDSEKQ 1598
OY 1403 -----YINDQGEN-----BKYLPF-----LNN 1420
DB 1599 DIKNFPPLFKINGNDMAFPYLPWLVPOQOQNDVNTLIQIRQLFOQOFMOQVRENELE 1658
OY 1421 IETLYKTIVNDKIDLVHLEKAVLNTYKESNVYKIKELNYLKTIDQKLADFKNNNFV 1460
DB 1659 LRQORNLLEKKLID--QIOLESQNLNNKQSEFSKVSAMEKL--LEKTESRLNDFQOKINY- 1713
OY 1481 GIADLSTDYNNHN 1493
DB 1714 -----LTKKVNQHN 1722


```
RESULT 45
T14867
Interaptin - slime mold (Dictyostelium discoideum)
C.Species: Dictyostelium discoideum
C.Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C.Accession: T14867
R.Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A.Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dicty-
os.
A.Reference number: 218248; MUID:98365468; PMID:9700162
A.Accession: T14867
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1738 <RIV>
A.Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC34582.1
C.Genetics:
A.Gene: abpd
A.Introns: 173/2; 1680/1

Query Match 5.2%; Score 438; DB 2; Length 1738;
Best Local Similarity 21.0%; Pred. No. 1.7e-06;
Matches 302; Conservative 254; Mismatches 528; Indels 352; Gaps 65;

QY 111 TNPSSSDAKSYADLKHVRNLYLTTELKYPOLFDLTNMLTLCNIGHFKYLIDG 170
DB 425 TNLQDLTBOO-QOYQESLKLVLNLELOE-KSNQLVKSNQLSM----- 469
QY 171 YEEINELLYKL-NFYDLRLAKLVNDYCOIPFNLRIRA-----NELDVLK 218
DB 470 QATNSLMAKIGLMDLNDIPIDIKED--EIANLKIESKNLKCPODDNALQSR 527
QY 219 KLVEGYRKL-DNIKDNVGMEDYIKKNTIENINELIEESKRTIDKKNKATKEEKK 277
DB 528 SLTIEGTOLDRKQILNELERDRDKFIEFTSSNQLADNQLDQLTN---EKQSI 583
QY 278 LVQAQVDSLTKQLE-EAHNLISVLEKRDPLAKKNENIKELDKINEIKNPPASGNT 336
DB 584 TLOLOQODIKERQEFQEQQLS---QIDST--TTNIOEQDKFNNDL---QEFNT 632
QY 337 PMLT-----LDKKKLIEHEKEIKELAKTIFNIDSLFDPLELEYLR 380
DB 633 QQTINQOETHRLTQQLXQINTDYNKQLOLSIKD-NQITN-----EQLKQSL 661
QY 381 EKNKNIDISAKVETKESTENEPNGVYPLSYND---INNALNE--LNS-FGDLIN 431
DB 682 EKDKETE---KLSNOQEOODEKINNMLLEIKERKDLIERINQOOLENIDLNKYOQLL 738
QY 432 PFDYTPKSPKNITDNERKKFIEKEK-----IKIEKK---IESDKSYEDR 477
DB 739 EPEENF-----LNSSEKEQNLNLOSKODERFNQINDELEKEKQSLIEDFNOYKQ 733
QY 478 SKSLN-DITKEYEKL---LNEIYDSFNNNIDLTNFKMKMGKRYKVEKLTNNFASY 533
DB 794 QLSNSNINQOQLOSTIIEISELKEQKELNDSKILIEKKOL-QOLOQEFQDLNKN---Q 848
QY 534 ENSKHNLEKLTALKTME-----DYSLNIVVEKELKYKNTL---ISKIENET 578
DB 849 KIHODOLELEKQLOLOQEDYDOLNETNOSIENQNLNKNENLNKEQOELIKQNL 908
QY 579 EPLVENIKKDEQLPEKKTITKDE--NKRPKEILEV-----SDIVKY 617
DB 909 NOQIEITQDPQOEFKONSINIEVLNKEKLTLOLOQDYDOLKQONSNDKEKNDLEK 968
QY 618 OVQKVLNKKIDELKKTOLILKNVELKNIHVPNSYKOENKORPYULIVLKEEIDKLKY 677
DB 969 ENQKLSIQNELN---QLIEKN-ESDHK-----EQQKQOSIENDLEKENQ----- 1010
QY 678 MKPVESLINEEKNKNTIEGQSDNSEPSTEGEIT---GQATTKPGQAGSALEGD--- 728
DB 1011 IOQLQSLNEQRO-----QOSNQLSEKQOQLNQLIEKNQFDQEQQLKQOSIENDLEK 1064
```

```
QY 729 -----SYQAQAEQKQAPPPVPEAKQVTPPAPVNNKTEN----- 768
DB 1065 ENQIQLOSLNBOROO-----SNQSEKQDQQLNQLIEKNESQKEQQLKQOST 1114
QY 769 -----VSKLDLEKLYEFLNTSYCHKYIIVSHSTNNEKILKOYKITEESKLS- 820
DB 1115 ENDLIERKNQLOLOLNE---QROLOSEVSIIDDKILELEKQKQOSDLKLINDEK 1170
QY 821 -----DPLDLEFNIONNI PVVYSPFDSLNSLSQLEMEIYEKEMVCNLVYL 866
DB 1171 QOODKQLODKQIBFDQLOLFFNQFN--DKDSQFILOLQDDQKQLOSLIOOD---LNQLKQ 1225
QY 867 KQNDKIKNLLEKAKKYSTSVKTLSSSMQSLSLTPDK-PEVANDQTHS---TNLN--- 920
DB 1226 ENQEKERQLEKDEK-----LOSIOFENQEKERQLEKDEKQSLIOQNLQNLN 1273
QY 921 --NSLKFENILSLGNKNITVQELIGQSSSENFYEKILKSDPFYNESEFTNFKSRADI 978
DB 1274 DEMQEKVKQFSEKDEKQSLQDQDLNQLKQENQEKQSLSEK-----EKQSLIOQDL 1325
QY 979 NSLND--SKRKLIEDINKLKTQLSFDLYNKKYKILERLFDKKRYGKYMQIKLT 1036
DB 1326 NQNLNDQIKKNEKLEKE---KEEQDLKQODPNQOQSLQLEKLESEKENQLOQLKQ-- 1380
QY 1037 LKEQLESKNSLNPK---HVLQNFVFPNKKRKEAIEAETNTLTKILKHHYGLVK 1093
DB 1381 -----ENETNLNQOQSNELITQQLKQQLKQOQOQNNNEKETER-LIOEIDQKQ 1433
QY 1094 YNGESSPLKTLSEESIOTEDNYASLENFVLSKLEGLKLDNMLE--KKKLSYLSGLH 1451
DB 1434 QQELDQSELN-KEIKIQ-----TQOEFQSL--NHRMSQDLHLQLOQELDQKQSPD 1485
QY 1152 HLIAELKEVIKKNKYGNSPENNVDYNNALSEY--KKFLPEGTDAVTVSESGDTLEQ 1209
DB 1486 DQDHQFKVY-DERYNLOLOLEQSTLSNNQDLKLEKLP-----ELDS 1530
QY 1210 SOKPKPASTHVAGSNMTTTSQNVDEVDVITVPIGSEEDYDGLGVATGEAVTPSV 1269
DB 1531 NEKQKITDILSNISNQLSLQNDKD-----LSEKNNSIKTLESITQQLSLIDE 1581
QY 1270 IDNLSIKENEVLYLKLPLAGVYRSKLQLENNVMT-----FNVVNDKILNSFPNKR 1323
DB 1582 KDWLINDLQKQKQOQPPASSPSSPSSPSTPTPKPQRNQIHLRLVMEIYVRNQ 1641
QY 1324 NFKNVLESLLPKDLTSSNYVVKDPKFLNKEKRDKFLSSYV---IKSDITDI 1376
DB 1642 -----DLIR-----KNKTKFYKLENGDIYVNSIYRSLSDDNDSDL 1678

RESULT 46
P00120
Major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (isolate B
N:Alternate names: gp195, PwMSA, PSA
C.Species: Plasmodium falciparum
C.Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Jun-2000
C.Accession: P00120
R.Kimura, E.; Matzel, D.; di Santti, S.M.; Scherf, A.
Gene 91, 57-62, 1990
A>Title: Genetic diversity in the major merozoite surface antigen of Plasmodium falci-
A.Reference number: P00120; MUID:90382698; PMID:2205540
A.Accession: P00120
A.Molecule type: DNA
A.Residues: 1-144 <KIM>
A.Cross-references: GB:M32111; NID:g160490; PIDN:AAA29701.1; PID:g552211
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; surface antigen

Query Match 5.2%; Score 435.5; DB 2; Length 144;
Best Local Similarity 65.4%; Pred. No. 1.2e-07;
Matches 100; Conservative 8; Mismatches 24; Indels 21; Gaps 5;

QY 30 KKLAEADAVLTGYSLFQKEKYNVLENGTSGTAVTSTPGSKGVSAGSGSGS--VASGGSV 88
```



```

Db      1 KKLLEDAVLTGCLFPEKKEMLNE-----BEITL-----KGASAGSSSGSPSPSG 51
OY      89 ASGGSVASGGSVASGGSNSR-----RTN-----PSDSSSDSDAKYADLKHRYRNL 137
Db      52 PSGTSGPSTGSPGCTSPSPSSRNTLPRNSNTSSGASPPADASDSDAKYADLKHRYRNL 111
OY      138 TIKELKYPQLFDLTNHHMLTLCDNHNGFKYLID 170
Db      112 TIKELKYPELFDLTNHHMLTLCDNHNGFKYLID 144

RESULT 47
JC6009
surface-located membrane protein Imp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JG6009
R:RefSeq: S.A.: Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis Imp gene system
A:Reference number: JG6009; MUID:9621016; PMID:8631664
A:Accession: JG6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CA64858.1; PID:g1197336
A:Genetics:
A:Gene: Imp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
F:957-992/Domain: tetratricopeptide repeat homology <TT1>
F:993-1026/Domain: tetratricopeptide repeat homology <TT2>
F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>
F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match          5.2%; Score 435; DB 1; Length 1302;
Best Local Similarity 20.3%; Pred. No. 1.5e-06;
Matches 317; Conservative 255; Mismatches 530; Indels 458; Gaps 70;

OY      118 SPSDAKSYADLKHRYRNLTIKELKYPQLFDLTNHHMLTLCDNHNGFKYLIDGYEINEL 177
Db      30 SNNSTKTKNEASRLN--LTKL-----DN-----EGQKANDF 62
OY      178 LYKLFFYFDLRAKLNDYCANDCQIPNLIKIRANELDYKLVEGYRPLDNI-KDNGV 236
Db      63 IAROD-----KKFNSTAFKNHS---NTSKLDEIFEFISK-----KILENLQKDEQ 105
OY      237 KMEYIKKNKTKIENINELIESKKTIDKNKNATDEEEKKILYQAYLDISTYNNKOLEEH 296
Db      106 RLEENLKNLKRFLD--LQNLINS-----NDGQVNDSSDAKALINENOIDSL----- 150
OY      297 NLISVLEKRIDTLK-NENI---KELLDKINEIKNPSPANGNPNTLIDKNKKEEHE 351
Db      151 -----PIDIKTKTNELEAKKELLKINAHEKL-----QSKIFNEKK 188
OY      352 KEIKEL-----AKTIKFNIDSLFTDPLELEYIREKNKNIDI-----SAKVE 393
Db      189 QELKRVLLDEDTKEVDFTKEQVFETETININETSIEDI--KNKIIIEKATSSLSKTL 245
OY      394 TKESTREPNEPYGAVYPL-----SYNDI--NNALNELNSGDDLINPDYIKPEPK 441
Db      246 NTKQDELQEFEN-IKKDLQDFINTKLNDAKYOSIKQAKLDKNSLNGI-----NK 294
OY      442 NITYTNERKKFNEI---KEKIKIEKKKIESDKSYEDRSKSLNDITKYEYKTL-----NE 494
Db      295 N-STIKELKAGNALIKAAKEAGLEKEL--DQGNIKDTLKTETINNAKREFKLLDNDQK 351
OY      495 IYDSKFPNNNIDLTNEKKAMGKRYSTKVEKLTHTNTPASTENSKHNLEKLT-----TKALKY 549
Db      352 IYDLSKNSLNDISKEQOISL-----DKESMESANDLNLKLEY 391

```

Qy	550	MEBYSLENIYAEKELY	-----YKLUKISENIE	-----TLENIKIKDEPELPEKI	597	
Db	392	KE--ILKFKOEBEAFNELEBOTRKHIEF	LDIEVKANNPNTATVLKDTLNKKA--	KKS	447	
Qy	598	TKDENKPEDEKILEVSDIV--	KVOVKVLLMNK-----	IDELKTYOLLKKNELKHNIHP	650	
Db	448	TJNSNKN-----	SDIIANELLQALADANKAKDOVEANK--		482	
Qy	651	NSYGOEKOEPPYLYIYKKEIDKLFKMFKYS	LNEEKNKTKTEGOSDNSEPEGEIT		710	
Db	483	-SKEQ-----	-LNLIDANTLPLQ-----	LNDNDELVYAKESLAEITNANKAV	527	
Qy	711	GOATFKEGQAGSALBGDSVOAQOQ-----	EOKQOPVPVPEAKAOVPTPAPVNNK		765	
Db	528	NQNNMAMOSAKSSLDOKVTYKIGNQUTERNKOKDA-----	KFELEQTRKIDIDNFL		578	
Qy	766	TENYSKLDYLEKYEPLNTSYICHAKIILV	SHSTMEKILQYKTKKEESKLSGCDPLD		825	
Db	579	TDDVKNPNPATVLVKDLYTNAKDKKSVTKS--	SNKSEIITAADELQALDAKKAVAKDID-		636	
Qy	826	LFNTQNNIPWYMGFOSLNNLSLSQLEMEIYEKEMVCN	ITKLEKDNKIKNLLEBAKYSTS		885	
Db	637	---EANKSKEQBSDSITMA-NOL-----	LKNLYDSK-----	DQAKTE	673	
Qy	886	VKTLSSSSMOPLSLTPDQKREVSANDTSHSTNLNS	SLKLFENILSLCKNNI-YOELLIG		944	
Db	674	LSOEIQASQOELMNNPSTSMOSAKESLDAKVEITRKLETF-----	NKRDVYKFKEL--		725	
Qy	945	OKSSENEYKILDSOTFVESEFTN-----	FYKSKADIDNSLDSKRRKLEEDINK		996	
Db	726	-----EKTRKIDIEFITMKTNTNNTSYLSEL	TSKDSKSNSTYNSNKSNDIETANTE		777	
Qy	997	LKKTLL-----	QLSFDLYNNKYKLIKLERLPDKKRYGKYMOKK--		1034	
Db	778	LKQALAKANTDKAQADNLA	STKEQNLNNSISSANFL-LAKLTDKDNITQOAKTELEKEVQ		836	
Qy	1035	-----LTLKEDESKLNSLNPKPHVLDNF	SVFPNNKKKEKELIETENTLE--		1079	
Db	837	KANOAVASNNTASQASAKSSLDKAVEITRKLETF-----	FNKKRDVKEFELDTRKDI		889	
Qy	1080	---NTRKILLKHKYGLVKYKNGSSSEPLKTLSEBSI	OTEDNVA	SLNFPVYSL--LEGRKL	11333	
Db	890	DEFINTMKTNPDSYTLSEL	TSKRDSKNSITJNSNKSNDIETANTE	LQALAKANTDKQA	949	
Qy	1134	DNL--NLEKKLISTYSGCLHHLI	ALKEVLKKNKNTYGSSENNTPYNNAL	LESYKFKLEPEG	1192	
Db	950	DNLARSTREQNLNNSISSANFLAKL-----	TKDWNITQOAKTELEKE		991	
Qy	1193	TDVATVYSESGDPLEOSQKRPASTHVAESNT	ITTSQNDVDEVDYIIVIFEGESEED		12522	
Db	992	VOKANOAVASNNTRASMOS-----	AKSSLDKAVEITRKLETFERNKODV--	KFELEQ	1042	
Qy	1253	YDDLGVVYTGAVTP--	SYIDNILSKIENEYEVILKPLAGVYNSL	KKOLENNVPNPV	1309	
Db	1043	RKDIIDEFITMKTNPNTSYLSEL	TSKRDSKNSI-----	TJNSNKSNDIETANT	1090	
Qy	1310	NVKDILN-----	SRFNKRBNFKVLESDILPYKDLTSSNYV	KDQYKPLNE--	1356	
Db	1091	ELKQALAKANTDKAQADNLA	RSTKEQNLNNSISSANFLAK--	LTDKDNITQOAKTELEKEV	1149	
Qy	1357	-KROKFLSSYNYI-----	KOSIDTIDINFANDVLYGYKILSE---	KYKSDLSIKRYIND	1406	
Db	1150	OKANOAVASNNTVSMOSAKSSLDOTKYVTEIT	KRKLETFERNKOKBAKFNDELKKTGRQO	IDEFINT	1209	
Qy	1407	KOGENEKYLPFLNNIEFLYTVN-----	DKIDLEVILHLEAKVNLTYTEKS	AVEYK--IK	1458	
Db	1210	NK--NPNVSELISQUTSKRPSKNSVTJNSNKS	DI-----	ETANTLEKQALA	1254	
Qy	1459	ELANTYKTIQOKLADFKKNNNPNVGIAD	LSDYNNHNNLITKLEFSLGWFENI	LATVATVSNLD	1518	
Db	1255	KANTDKAQADNLA	RSTKEQ-----	LKNSISS-----	ANTLAKLTD	1250


```

Db 2224 QLLATLSKIINKLN-----VKVSELDH-DEFSYVLLLSQKSKMNTIEFNKEKE----- 2269
Qy 1517 LDGNOGLMNTISOHCYKCKOCPQ 1539
Db 2270 ---QLSKILKTSIHLILNSIDIQ 2289

RESULT 49
T30822
Imp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect Immun 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis P621 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882; PMID:7543881
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U01962; NID:9790243; PID:9790244; PIDN:AAA0103.1
C:Genetics:
A:Gene: Imp1
A:Genetic code: SGC3

Query Match 5.1%; Score 433.5; DB 2; Length 1365;
Best Local Similarity 21.1%; Pred. No. 1.8e-06;
Matches 318; Conservative 251; Mismatches 530; Indels 411; Gaps 74;

Qy 129 KHRV-----RNYLLTKELKYPOLFDITNMLTLCDNHSGFKYLIDGEEIN----- 175
Db 33 RHKIANOSOKENLLQNEKKILO-KOLNELGYKYINEIWFHEQEVLOSLKINNKSETKA 91
Qy 176 --ELLYKLNFPDILRAKLNVCANDCOIPFNLIKIRANEIDVAKLVFGYRKLPLDNIKD 233
Db 92 IEESTLELKDAITLLISKIKN-----QINOKELPAK-----FNEIKD 129
Qy 234 NVGKMEYDIKK--NKKTIEINELIEES-----KKTIDKNKATKEEKKKLYOAQ 282
Db 130 ---KLOGYIKNELSKOEYEHKQNIENELNKYTPISLESTLIEIQNATNML----- 177
Qy 283 YDLSTYKOLEAHNLISVLEKRIDTLKKNINELDKINELKINPPRANGNPNTLLD 342
Db 178 --IKLNESTRKEDN-----IDNLNAKQOLKASISQAQQL--POLSDN--DSEIA 222
Qy 343 KKKKIEHEKEIKELAKTI--KENIDSLFTDPLELEYLRKKNKNDISAKVET-----KES 397
Db 223 KAKK--SLDAEIKNANOAVASNNTASMOASAKSLD-----AKVAETIKKLETPKKDEA 274
Qy 398 TEPNENGVYTPPLSYNDINNALNELNSF-----GDILNPFDTKEPSKNITYDN 447
Db 275 -----KPNELKQTRNOIOEFINTKNNNPYSELLISQLT--SKDSKNSVYDS 319
Qy 448 ERKFFIEIKETIKIEKKIESDKKSTEDRSK-----LNDITKEKELNIIYISKFPNN 503
Db 320 SNKSDISANTELKQALAKANADVQADNLKSIKEQLNVSANNTLSAKLTJOK--DNT 377
Qy 504 ID--LTNFEKMMGKRYSYKVELKTHNHTFASYSKHNLE-KLTKALKYMEDYSLRNIVV 560
Db 378 IQAKTELEKTV-----QKADQAIKSNNTASMOASAKSLDKAKVAELTKKLETFN-----K 427
Qy 561 EKELKYKKNLSIKIENETIETLVENIKKDEOLFEEKITKDKENKPEKILLEVSDIVKYOVQ 620
Db 428 DKEKKF--NELKQTRNOIOEFINTKNNNPN--YSELISQLTSKDSK----- 470
Qy 621 KVLAMNTI-DEKKTQULILKAVELKHNHVPNSYKQENKQEPYLLIYLKKEIDKLKVPMP 679
Db 471 ---NSVTSDSNKSDIESANTEKQ-----ALAKANADKV----- 501
Qy 680 KVESLINEKKKNITBEOOSNSEPSTEGEITGQATTKPG--QOAGSALESGSQAQAOEQ 737

```

```

Db 502 QADNLAKSIKEQL-----NNSVSNANTLSAKLTDKDNTIQOAKTELEKE-VQKADQAI 553
Qy 738 KOAQPPVVPVPEAKAQVPPAPVNNKTEVNSKIDYLEKIEYELNYSICHKYLIVSHS 797
Db 554 KSNV---TASQASAKSSIDAQVAELTKKLETFNK--DKAKFNEIKQTRNOIOEFINTKN 609
Qy 798 TMM-EKILQYKITYKEESKLSQDPLDLFNQNNIPVMSMPSLNSISQLEMEI-Y 855
Db 610 NPNVSELLSQ--LTSKRDSKNSVTDSSN-----KSDIESANTEKQALAKANA 655
Qy 856 EKEVNCNLYK-LKD--NDKIKN-----LLEAKKYSTSVKTLSS 891
Db 656 DKVADNLAKSIKEQLNVSANNTLSAKLTDKDNTIQOAKTELEKEIQKANOAIKSNNT 715
Qy 892 SSMOPL--SLTPQ-----DKPEVSANDTSHSTNLSNLSKLEFNLSIGKKNKINYOELIG 944
Db 716 ASMOASAKSSLAQVAELTKKLETFNKDEAFNELKQTRNOIOEFINTKNPNVSELS 775
Qy 945 QKSENEFYEKILKSDTFYNESTFNFKSKADINDSLNDESKRRKLEEDIKLKTQLS 1004
Db 776 Q-----LTSKRDSKNSVTDSSNKSQSDIESANTEKQAL--- 807
Qy 1005 FDLNKKYKLEKLEFDKKKTYGKKMQIKITLKEOLESLNLSNPKHVLQNFVPEFN 1064
Db 808 ---NTAKA-----KSSID-----NELRPLKNDLOSKIEEGPIRNT--NFSWISS 848
Qy 1065 K-----KKEAIEAETE--NTLENTKILKHYGLVAYNGESSPLKTLSEE-SIQTE 1113
Db 849 KLETTKNNLAELRKADIKKNNPSSQALDKSSQOVQKLNEL--LKTIEEGKVETK 906
Qy 1114 DNYSASLENFKY-----LSKLEGLKDONLLEKK-----LSYSSGLHH 1152
Db 907 NSNIGYRLFKLAQAEQFNNSDVKLKNAMEKQTLKSKOKRLGNOSTKDYLTOLSTEMST 966
Qy 1153 LIAELKEYIKKNKNTGNSPENNNDVNNALSEYKFFLEPGTDAVAVSESGSDTLEQSQP 1212
Db 967 QESTIKKAYIVNI---QAHIRNNLSQYRLADKLI-----ANM 1001
Qy 1213 KKPASTHGAESNNTITTSQNVDEVDYIIVIFGESEEDYDGLQVYTG-AVTPSYID 1271
Db 1002 KRQYGDVKIGES-----LOQMOLDMDSDVL--SYDSTLKDDEPKALRYLVGQTRKPPVSS 1055
Qy 1272 NILSK-----IEN-----EYEVLYKPLAGVYSKQLENNVMTFNVNV----- 1311
Db 1056 WFINRNSIENYONLRNLLIVRENEILLDKA-----KDLKRAKTIKYEVDENINSIDQ 1110
Qy 1312 -----KDLNSRPNKRENFKNVLESDLIPYKDLTSSNYVVKDPYKFLN-----KERR 1358
Db 1111 RAKRLKQELIINAK--NDLSNFTLNHQKNQFTAKDITPKISLEKNKLNELNOYLLPIIKKA 1169
Qy 1359 DKPLSSYNYIKSDIDTDFINPANDVLGYKKILSEKYSQSDLSIKKYINDKQEGENEKYLPL 1418
Db 1170 VSKISEIEKNKEKLE-DIIRSENFYL-WKVEINKYISEL-----TNKQVELSRNINFE 1220
Qy 1419 NNIETLYKTV--NDKIDLFYHLEAKVNLVYTEKSNVEKIKELNLYKTIOQ--KLADKK 1475
Db 1221 NKWESIKQTDLNLNKENVSLKLEYIIN-----NSNAQYSINRI--LSTVPFIFIAVQOTR 1274
Qy 1476 NNFVGIADL 1485
Db 1275 SNNLRSLAEI 1284

RESULT 50
E71606
hypothetical protein PR0765v - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71606
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.;
Perita, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

```


QY 512 MNGKRSYKVEKLTHTNTFASYSKSHNLEKLTALKYMEDYSLRNIVVEKELKYKYL 571
 Db 1313 LMKKKKQSPDKEL-----NSLKAELEALFKA-----KSVESKKNDSENEK 1353
 QY 572 SKIENELETIVENIKKDEOLFEEKITKDEKPKILEVSDYKVOYVYLKMKIDEL 631
 Db 1354 AALSEEDIDANNEKIKNQALD--RKATADLOEANEKRAE--VEAORDKILVADN--KM 1405
 QY 632 KKTOLIKNVELKHNIHVPSYSKOENKQEPYLLVLKKEIDKLVFMPKVESLNEKNK 691
 Db 1406 TKI---LEETIKADE---EMTYKVEN---YKVLKRRKAD-----LEENANEN 1443
 QY 692 IKTEGOSDNSEPTGEITGQATTKPGQAGSALEGSVOAQAQEQKQAPVPVPEA 751
 Db 1444 LDIEKKRMMKKEKQVKKLEGELETKDKLNAALAEKDSI--FTAKKQSDAD----- 1492
 QY 752 KQVPPPPAVNNKTEVNSKLDYLEKLYEFLNTSYICHKYLIVSHSHTMNEKILKQYKIT 811
 Db 1493 -----LEELNKTEVEHDEVAANKLTQITKLTNDQSAE 1525
 QY 812 EESKLS-----CDPLDLFNIONNIPVYGMFDSLNLSQLF 851
 Db 1526 EELNELRSKADKKKKISELEOVNELESRPVGTGNADENEIKTRDQIADLNKALEMG 1585
 QY 852 MEIYEKEMVCNLKYLKND---KIKNLEEKAVSVTSKTLSSSQWPLSTFPODKPEVS 908
 Db 1586 VQNNQATATKKEKAKDNDSITKIEITENEMKKTLENKAKRLE-----ODKDEA- 1633
 QY 909 ANDTSHSTLNNSLKAFENILSGKKNKIYQELIGKSSSENFYEKILKOSDTPYNESFT 968
 Db 1634 -----DKAVSQITKRRGLEEYVKKLTTEQIA----- 1660
 QY 969 NFVSKADINSINDESKRRKLEEDINKLKTQLSFDLYNKYKLEBRLFDKKRYGVKY 1028
 Db 1661 --LKFOINAPSVAGQEEKQRLSEDIAMK-----EQLEQERTTANA 1701
 QY 1029 KMQIKTLLKBOLESKLSLNPKHYLVNFVFPNKKKEALFETENTLNTYILKHY 1088
 Db 1702 EARRKKIQAELEDEYKFLNEVNTQREKL--VAKNSEDAEI--DSLKEEKKALDEEI 1754
 QY 1089 KGLVKNYNGESSPLKLTSEESIQTEDNYVASLENK-----VLSTKEGLK--KNLNLEK 1140
 Db 1755 EKITDNN-----KLSIEDSLDKRYNALLDSKDSVSKKEKFOBELKATYKTALETEK 1807
 QY 1141 KKLSTYSSGLHLIAELKEVIKKNKNTYGNPSFENNTVNNALESYKFLPEGTDVATVVS 1200
 Db 1808 KN-----HAETMRKGRLE-----KEAAEVQVRLALQKNL--DLA---- 1841
 QY 1201 ESGSDTLEQOPKRPASTHYGASNTITTSQNVDEVDVY--IVPIFGESSEEDYDLGQ 1258
 Db 1842 -----OQERAKAKRKYRAADGELKSLMNELDVKKDLDKQADDLADKEDLALDQ 1892
 QY 1259 VVTGEAVTPGVINDILSKIENEYEVLYLKPLAG-----VYRSLKCOLENNVTMFNV 1309
 Db 1893 KYTLYKQKQSVFD---SRIDMQOQLDLEK--AGRAKAKQOKQAYEKKLOELQENDNDFE- 1947
 QY 1310 NVKDILNSRPKNRENFKNVLESDLIPYKDLTSSNYVVKDPYKLNKERRDFLS----- 1363
 Db 1948 EYKETAADKRIINTLSAQKODLOKELEKERGLKQSEKEVQRLRVACQOELTKTADYETKEASFNT 2067
 QY 1364 SYNTIKDSIDTDI---NFANDVLGYIKLISEKYSKSD---LDSITKYYIND----- 1406
 Db 2008 SIKVKAKYAEAEIEELTTEADALAKAKMAEKKAKTSQKIKDELQKTADYETKEASFNT 2067
 QY 1407 -----KQGENEKYLPFLNNITLYKTVDKID-----LFVHLEKAVLNTYEKSNVEVK 1456
 Db 2068 EIKGTQAEIKKYQOQVVDDETFRMSSTLEDIKGTALANQOLELDKVNKOYEKLLKQYK 2126

RESULT 52
 S48385
 hypothetical protein Y1149c - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae
 C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
 C:Accession: S48385
 R:Churchev, C.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: S48310
 A:Accession: S48385
 A:Molecule type: DNA
 A:Residues: 1-1679 <CHU>
 A:Cross-references: GB:247047; EMBL:Z38059; NID:g603997; PID:g763197; GSPDB:GN00009;
 A:Gene: SGD:MUP2; MIPS:Y1149C
 A:Cross-references: SGD:S0001411
 A:Map position: 9L
 Query Match
 Best Local Similarity 197%; Pred. No. 4.8e-06;
 Matches 335; Conservative 296; Mismatches 583; Indels 488; Gaps 78;
 QY 198 NDYQOIPPNLIRANELDVLKLVGYRK--PLDNKDNVGMKME--DYIKKN-KKTLEN 251
 Db 6 SEFLNVPEE-SLQGVTVPLVRKL---YKKIAKFERSEEVTKLVNVADEIKSQYRSRISK 61
 QY 252 INELIEESKTIIDKKNKATKEE-----EKKKIQAQVDSLTYNKQLEEAHNLISVL 302
 Db 62 LKQLLDES---SQKNTAKBELNGLDQOLNEERSRTREID--ALKQQLVSHBAMREV 115
 QY 303 --EKRI-----DPLKKNENIKELDKINETKNPPAN----- 332
 Db 116 NDEKRVVEEYDIWQSRQGNDSLNDLKNKELRRKLMEMENILQKRSKVAISLQLYD 175
 QY 333 -SGTPTNLTLDKKNKIEE-----HEKETELAKITKFINDSLFDPL 373
 Db 176 TSVOEKELMIOSKKLIBEEKSSFSKRTLTREVTSSHYENLEKLYQWQSVESVFT--- 232
 QY 374 ELEYLREKKNKNDISAKVETK-----ESTEPENYNGVTVPLSYNDINNA----- 419
 Db 233 ---YKFLNQNQLOSQVEKYLEMKNLKPASVEKAEFEKMTLQKNMNDLRSQULTS 289
 QY 420 -----LNLNLS-----FGDLINPFDYTK--EPSKN----- 442
 Db 290 LEKDCSLRAIKKNDNSCRNPEHTVIDELDTKRLREKSKNECORLQNIWMDCTKEBEA 349
 QY 443 -----LYTQNE--RKKFINEIKIKITIEKK--IESDKK-----SYEDRS 478
 Db 350 TMTTSAAVSPVIGKLFSDIKYKLRQILKERQKQFOLQNOLEDFILELEKRPBELLISFERT 409
 QY 479 KSLNDITREYKELKLENIYDSKFNNNIDLT-----NFEKMGKRSY-KVEKLTN 526
 Db 410 KSLHELRKSTLELETYSLTRKQREITSLRKINGCEANIHSLVQRDLARQVKKLL 469
 QY 527 HNTFASYEN-SKHNLKLTALKYMEDYSLRN-----IYVERELKYKKNLSKIENTET 580
 Db 470 LNTSAIGETASPLSQDELISRLKILLESNIYVENDSQAILTERLVSNNVMELEKNVEL 529
 QY 581 L-----VENIKKBOLEFEKKITKDEKPKILEVSDI--VKQVQVVLNKNID- 629
 Db 530 LNCIRILADKLENYEGQDKTLQVENQTIKEADAAILIEENINAKMETRINIILREDS 589
 QY 630 -----LAKTKQL-----ILKNVELKHNIHVNS 652
 Db 590 YKLLASTEENKANTNSVTSMEAREKKIRLELEALSTKVENSAITQILKRELLIYKQ 649
 QY 653 YKQENKQEPYLLI-----VLKKEIDKLKVMKPR-----VESLJNEKKNIKE-GQS 698
 Db 650 CKKTTLEDPEPENGLAKKEREMLEBALDHLKALEKQSMVPYIIVIEKRASTELSQS 709
 QY 699 DNSEPTGEEL-----TGQATYKPGQAGSALLEGDSVOAQAQEQKQAPVPV 746
 Db 710 RIKIKLSLEYEISKLKKEFASPIPTKRESLITRDFEQC-----CKEKKELQ-----M 753
 QY 747 PVPEAKAQVPPPPAVNNKTEVNSKLD-----YLEKLYEFLNTSYICHKYLIVSHSHTM 799


```

Db 754 RLKSEI-----SHNENKMFSSKEGQYAKIKELENNL-----787
Oy 800 NEKILQYKITEESKSLSCDPLLEFIONNIPVMSMPDLSLNSQLFMEIYEREM 859
Db 788 -ERLSDLOSQIOETJESTRSCDSOLK-AQNTI-----DDTEKMSLTLTSLNKET 838
Oy 860 VCNLYKLNDNDKIKMLLEAKKVTSTSVTL-----SSSMQPLSLTPQDKEVSANDTSH 915
Db 839 TIE-----KLSEIENLDKRELKRTKFOYKFLDQNSDASTLEPLRLKELEIOQLD-----890
Oy 916 STNLNLSLKLFEINILSLGNKNKIYQELIGQ--KSENFYEKI-LKSDTFYVESFTPEVK 972
Db 891 ---ANSOIAQYEIIS--SNEVALIELKNEIEMAKTENYDAKIELEKKEKMAEEDLSRL 945
Oy 973 SKADINSINDESKR-----KLEBDIKKLKKTQLOLSDLVNNKYKLKLERL-----1018
Db 946 GELGELRLQPLKGCALHPVOOSEKLRNEVERIQMIE-----KIEKSTVIQ 994
Oy 1019 FDKKTVGKYKMOIKR---LTLLEQLESKLSLNNPKHVLONFSVFENKKEAEIAETE 1075
Db 995 LCKKKEMSQYOSTKMEKNDLSFLVIRLEKDADC-----QAEILTKK 1036
Oy 1076 NTELETKILL-KHYGLVKKYVNGESSPLKLTSEESIQTEDNYASLENFVKSLKEGLKD 1134
Db 1037 SLYSAODLLDKHER--KMEEKADYERELISNIEQTES--LRYENSVLIEKVDOTAAN 1091
Oy 1135 NLNLEKKLSTYLSGLHLIAELK-----EYIKKNTGNSPSENNNDVNNALES 1184
Db 1092 NGDKHLVLVSFLMIRHERNSLETKLTCKRELAFVQKN--DSLEKTINDLQRTQTL 1148
Oy 1185 YKFFPEGTDTVATVVSSESDTLEOSOPKPASTHVGASNTI--TTSQVNDDEVDTI 1242
Db 1149 SEK---EVQCSAVIIDEFKDITKEVTO-----VNLKENNALLOKSLKNVTEKRE--I 1197
Oy 1243 VPICESEDDYDGLQ--VTGEAVTPSYDINILSKIENEYEVLYLKLPLAGYRSILKOL 1300
Db 1198 YKOLNDROBEISRLQRLDIQTKEQV--SINSNKKIIVYSEMEQCKOR---YQDLSQO 1250
Oy 1301 ENNVTFVNVVYKDI-----LNSRFRKRENFKNVLESDLIPKDLTSSVYVAKPYK 1351
Db 1231 K-----DAOKKDIKELTNEISDLGKLSAENANADLENKR-----NRLKQAEH 1295
Oy 1352 FLNKEKRDKF-----LSSYNTIKDSIDTDINFANDVLYGKILSEKYSDLISIKYIND 1406
Db 1296 KLDASKKQQAALTNELNELKAKDKLEODLHEFN-----AKYIDDTKKAHELOSEVDS 1350
Oy 1407 KOGENEKTLPLNNIETL-----KYIVNDKIDLFVHLEAKVLYYEKSN-VEVKIKE 1459
Db 1351 RDHEKDTYRTIMEEIESLKKELQIFKTANSSSDAF---EKLKVNMEKEDRIDERTKE 1406
Oy 1460 LNYLKTIDDKLADFKKNNNFVGIADLSTPDYNNNNLLTFLSGMVFENLAKTVLSNLDG 1519
Db 1407 --FEKKLOETL-----NKSTSSEAYSKD-----IETLKKMELKEYEDE 1443
Oy 1520 NLOGMLNTISOHCYKQCPONSQCFRHLDERECKCLNTYKOE-GDKCVENPN--TCNEN 1577
Db 1444 TLRIRKEAENLKRIRLPSERIRQIILSKRKE-----ELBEFRKKLKEMNGSLTFIDN 1498
Oy 1578 NG-GCDADAKCTEEDSGSNCK 1598
Db 1499 KSGGEDAEELMNSPCKGNSR 1520

```

RESULT 53

S41649
 DNA polymerase - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 20-May-1994 #sequence_revision 26-May-1995 #text_change 09-Jun-2000
 C:Accession: S41649
 R:White, J.H.; Kilbey, B.J.; de Vries, E.; Goman, M.; Alano, P.; Cheesman, S.; McAleese, Nucleic Acids Res. 21, 3643-3646, 1993
 A:Title: The gene encoding DNA polymerase alpha from Plasmodium falciparum.
 A:Reference number: S41649; MUID:93376482; PMID:8367280

A:Accession: S41649
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-1855 <WHI>
 A:Cross-references: EMBL:L18785
 C:Superfamily: Plasmodium falciparum DNA polymerase

Query Match 5.0%; Score 423; DB 2; Length 1855;
 Best Local Similarity 19.3%; Pred. No. 5.3e-06;
 Matches 385; Conservative 303; Mismatches 667; Indels 638; Gaps 88;

```

Oy 1 MKIIFLCSFLPF--IINTQCVTHESYQELVYKLEALEDAVLVGLSYLFOREKAVLNEGTS 58
Db 1 MYLIYLLKFFPFSLINDV-----DKEQVVRNQLD-----FLREKK-LKTLIS 47
Oy 59 GFAVYTTSTPGSKGSVASGSGSVASGGSVASGGSVASGGSVASGGSVRTNPSDSS 118
Db 48 GNLK-----SKEELHK-----TKNIEELMDQNTSS 72
Oy 119 DSDAKS-----YADLKH-----RVRYLLTIKELKYPQLEDLTNHMLTLC 158
Db 73 DSDGESRTKRRKRTRESYENSMYKNNKSSCYLRSKDYENIYDMKTNEYN-----123
Oy 159 DNIHGFYLDIGEYEEINELLYKLNFFYDILRAKANDVCANDYCOIPFNLIKIRANEDVLK 218
Db 124 -----KTIENNNNNN-----NYT-----YNNYNDMSFSNK-----DLEYSK 156
Oy 219 KLVFGYRKLPLNIKQNVGMEYIKKNNKKTIEININELIESKKTIDKNNKATKEEKKLU 278
Db 157 -----YTK-VKNEEDIDKHQPY-----DIHDMGSSLEKKENYVQEOQSNM 198
Oy 279 YQAVDLSLYNNQLEAHNLIS--VLEKRLDILKKNENIKELDKINELKNPPASQNT 336
Db 199 INVGNDDI-NEKKNHILAKNELVYKKEIQIKENENIEKVEQSGKNNNNKNNNDNN 257
Oy 337 PNTLID-----KNNKIEBEKEIRE-----356
Db 258 NNNNVDDFYNDYILNEMNISPLFIKIENTETLKEEKKLENDINDKEEDILENDISEML 317
Oy 357 -----IAKTIKFNIDSLFDPLELEY 378
Db 318 KKKYEQILKEEENNLNVVVYVFDICKHNSIILFEGTLTKKKYKYSISIFENIDRYYY 377
Oy 379 -LREKKN-----NIDISA-KVETKESTEPN 401
Db 378 FLNKKKYVYENGEEIKFNEKFKIHMSEFLBEFKIIEHYHNKAKAKYIKRKNLNS 437
Oy 402 EYPNGVYTPLSYNDINNALNELNSFGDLINPDTKEP-----SKN-----IYTONE 448
Db 438 SYDDELYIKVLVSYNNDPHEKFOKGSYLSFYCCNEDIVENFIKKNKFLPCWIKIKRL 497
Oy 449 RKKFLN-----ELKEKI-----KIEKKIESDKSYEDRSKSLNDITREYEKLIN 493
Db 498 RNDNSLNTLYCYEDCIVEDEKKSILLDCKVHEKRVKVDTSKIDQTSPIKINNVSNNNNNN 557
Oy 494 EYDSKFNNDLITNEPEKMKGRYSYKVEKLHNHNFAPYENSKHL-----EKL 543
Db 558 NNNNNNNNNNNNN-----NNNNNNNNSSNFGVYIISDQPKKI 600
Oy 544 TRALKYMEDYSLRNIVE-----KELKY-----567
Db 601 EGYQLSDINMLKLYIKVVSLLNEENVHEIFSICSLVQIDKLKYINFGISIRKSKKGT 660
Oy 568 -----KNLSIKIENE---IETLVENIKDEEOLF-----593
Db 661 MYNNKTTLCKLPDNEELLHTFLEKIKDDIDYIGYNTLNFDLEPLIRCHVHNLSYI 720
Oy 594 ---EKKITDEKRPPEKILVSDIVYOVQYVLNMLKIDELKKTQILKENVELKHINHP 650
Db 721 LSRKKKKKKNE-----KKAKYK---FNGINSTGSMNIIONINIGRLIVDIY 763
Oy 651 NSYKQENKQEPYLLVKKKEIDKLK-VFMPKVESLINEKKNI---KTGQSDNSEPSTE 706

```



```

Db 764 LKCRDSIKLTYY---CLDEIIDHYKKNFKODSKQVNTKNTSTNLNKTNTYNOOSVVSQ 820
QY 707 GEITGOATTRPGOAA---GSALEGDVSOAOAOBOKOAP-----VVPVPEAKAQ---- 754
Db 821 SVVSQOSV---SQOSVVSQOSVVSQOSVTSOSIOSQANPNHTNOSTHMHVTEEH 878
QY 755 -----VPPAPAVN-----NKTENVSKLDYLEKYLEFNTSYICRKYILVS--HS 797
Db 879 ASNKATGNHDIPTDVASTNSHRSNTHINIEF--DFLVNNINILJNCSTN--HLYDAQSIAN 936
QY 798 TMMKILKQVKITKEESKLSQCDPLDLFRIONNIPVMSMFPSLNSLSQLEMEIYEK 857
Db 937 NLNTVNSQJECINE---TVNVCNIIQI-----EKTRDLTKLGYIMWR 978
QY 858 EMVC-----NLVKLKDNDKIKNL-----BEAKKYST-----SVKTLSSSQMPL-- 897
Db 979 SLCTYSEVRFELHHEYNKKKFTPLIKRTRKIEIETNOKMKNTAFYLGSLVLDPLCG 1038
QY 898 -----SLTPQDKPEVSANDOTSHSTNLNSLKLEFNILSLGKNNIYQELIGQK 946
Db 1039 YDFEVLVDNFNSLTPSIIEIENVCFTLKLKNCVSIIE--DNKINTKKNKNI----- 1088
QY 947 SSENFEKILKSDPTFNESTNTFVSKADINDSLNDESKRKKLEEDINKTKTLOISFD 1006
Db 1089 --NNDTEKNI--HSDAEKNIHSDDKNTHSDCKNYDNDKSTYNKLEDD--NLENNVEI--D 1142
QY 1007 LYNKYK-----LKLERLFDK-----KKTGVKKYKMOIKKLTLLKEODESKLNSLNNPKHYLQ 1057
Db 1143 EFDPSKRGILPCLIKSLVEXKSVIKKLISNENKKEKELLIOGLSLIKLIS--NIIYGLG 1201
QY 1058 NFEVFFKKKKAE--IaETENTLENTKILKHYKGLVKKYNGESSPLKTSSESIOTEDN 1115
Db 1202 NTNRFKAKHIASTYTSKGRLLQHTKFKVEKFEFNKLVIIYODTDSIM-----IDTGIK 1254
QY 1116 YASLENKVKLSLEGGKLEKDNLEKKKLSYLSGLHHLIAELKVKINKNYNTGNSPSENN 1175
Db 1255 ANNTNNKESFKLAHLIKNSINKNYKKLEL---DLECFPSKL--LKKKTKACKAKVIDNN 1310
QY 1176 TDVNNNALESYK-----KFLPE-----GTDVATVVSSESGDSTLEOSOP---KRPAS 1217
Db 1311 -----LEKYEYEMKGINFKRDSKISKILGNELVRIIF--TNNDVOSKNIPVLENDLS 1363
QY 1218 THWGAESNTTTS--QANVDEVDYIIVPIGESEEDYD--LGOVYTGAVPSVID-- 1271
Db 1364 EQLHEVLTJNORIONDEFLDYIITKKLKNVHEYODKNSLGHVLAERH--IKDGY 1420
QY 1272 NIIISKIENEVLYLKRLAGYRSLKQOLENNVTFENVNVKDIINSRPNKRNENKYNLES 1331
Db 1421 NTCVNNKEIQCVCSTSEDASRFYKTSERKLNSNOCCEFSINEIKTYNLKIDKEYYIRNOJLS 1480
QY 1332 D---LIPYKDLTSSN-----YVVKDPYKFLNKKERDKFLSSYNYIKDSI-----DTD 1375
Db 1481 PINRLOCIIGTSAKLSGCCFNIVDKE-----IKTDQIEE--NLETFVLSLINESD 1532
QY 1376 INPANDVLGYKILSEKYSKSDLSIKKYINDKQENENEKYLEPFLNNIETLYKTVNDKIDLF 1435
Db 1533 ERFRDILKGFIVCS---KCMHNVCNPIETFKYFKCKCLTYLS---IDQIRNYIFSF 1583
QY 1436 VIHLEAVLVYTKESKSVNEYKIKELNLTQIQLADLQKKNVNVGIADISTDNNHNL- 1484
Db 1584 IHHLCNTFYKQLYICOGCTLTKTRI-----FLKNDK--NCPNINCEYTNKNSLK 1629
QY 1495 -----LTFELSTGWAFENILAKTVLSNLDGNLQGLN-----ISOHQCV 1533
Db 1630 PLISKKIYILILEYFLF--LLKDNL--KIPSNLVEXKKSSEDLNLTQOTNNENDITSNQE 1686
QY 1534 KKQCPQNSGCFRHLDERECKCLIN---YKQEGDKCVENPNPVCNENNGCGDADAKTEE 1590
Db 1687 EAKDPNGSISINHDINDNEHSEKNNHNIYK--NNSNNNNNNNNKDIENEYCND- 1739
QY 1591 DSGNGKKITCEG 1603
Db 1740 -----FIVCLC 1745

```

```

RESULT 54
S46773
myosin heavy chain - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YHR023w, type II myosin
C:Species: Saccharomyces cerevisiae
C>Date: 07-Oct-1994 #sequence_revision 11-Aug-1995 #text_change 23-Mar-2001
C:Accession: S46773; S05806; S12323
R:Du, Z.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 8082.
A:Reference number: S46773
A:Accession: S46773
A:Molecule type: DNA
A:Residues: 1-1928 <DU>
A:Cross-references: EMBL:U0399; NID:9500689; PIDN:AAB68872.1; PID:9500691; MIPS:YHR0
R:Watts, F.Z.; Shields, G.; Orr, E.
EMBO J. 6, 3499-3505, 1987
A:Title: The yeast MYO1 gene encoding a myosin-like protein required for cell division
A:Reference number: S05806; MUID:88111539; PMID:3322809
A:Accession: S05806
A:Molecule type: DNA
A:Residues: 1-45, 'T', 47-58, 'S', 60-329, 331-342, 'S', 344-528, 'ARGHGR', 536-540, 'V', 542-59
A:Cross-references: EMBL:X06187; NID:94015; PIDN:CAA29550.1; PID:94016
R:Sweeney, F.P.; Watts, F.Z.; Pocklington, M.J.; Orr, E.
Nucleic Acids Res. 18, 7147, 1990
A:Title: The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide sequenc
A:Reference number: S12323; MUID:91088308; PMID:2263482
A:Accession: S12323
A>Status: translation not shown; significant sequence differences
A:Molecule type: DNA
A:Cross-references: EMBL:X53947
C:Genetics:
A:Gene: SGD:MYO1
A:Cross-references: SGD:S0001065; MIPS:YHR023w
A:Map position: 8R
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; nucleotide binding; P-loop
F:78-779/Domain: myosin motor domain homology <MYOT>
F:180-187/Region: nucleotide-binding motif A (P-loop)
F:459-528/Domain: actin binding #status predicted <ACBP>
F:166/Binding site: ATP (Lys) #status predicted
Query Match 5.0%; Score 421.5; DB 2; Length 1928;
Best Local Similarity 20.8%; Pred. No. 6.2e-06;
Matches 304; Conservative 246; Mismatches 487; Indels 425; Gaps 73;
QY 121 DAKSYADLKHVRNYLLTIR-----ELKYPOLFDLTNHLTLCDNIGHGFKYLIDGYEIN 175
Db 765 DTKVY-----KIGMTKLFKAGVLADLEKQKDVKLNNIMIKLTFTIR-----GYTVRK 812
QY 176 ELTKLNFYFDLAKLNDVCANDY-----CGIP--FNLRKIR-----ANELDVLKLL 220
Db 813 EITVHLO-----KLTGTRVIGNTFRLYKNRLVKEDPWNELTRIKPLTLTSSNDMTRPK- 865
QY 221 VFGRKPLDNKIDNVGWMEDYIK-----KNNKTI---ENINELIEESKTTIDKNN--ATKE 272
Db 866 ---FNEQINKKLNQLOEMESKKKFLPEKNQKTVALENTQDLOLNQENENLKNSSLNRV 922
QY 273 EEKKKLYQAOYDLSTYK-----QLEBAHNLISVLEKRI---DTLKK----- 311
Db 923 KTSSETLOKOFDDLVSEKDEISREKLEVAONLEBAHOKIQLOETIREREATELEKLSKN 982
QY 312 NENIKELDLKINELKNPFPANSGMTPTLLDKN--KIEEHKEKELAKTIKFNIDSIFT 370
Db 983 NELTKQISLQNDLISKQSSOS-----LIKESKLKLENETIKRLKDVYNSKEEIKS- 1035
QY 371 DPLELEYLAEKNNKNDISAKVETKESTEPNEYNGVYTPSYNDINNALLNELNSF----- 426
Db 1036 DKL-----SSSEEDLDIKLVTLF-----KNCNIAMSRLOSLVTEN 1070
QY 427 GDLINPFDYTKEPSKNITYDNERRKFTNEIKK-----IKIEKKKIESDKSYEDRSKSLN 482

```


Db	1071	SDL	-----	BSKNBNPFKKEAALNOLNKNESSLK	-	KEKIDNKKELATPSKORD	1120
Oy	483	DITVEYKLLNEIYDS	FNNNIDL	TNE	EKKMKRYSYKVEKLT	HNNFPASYENSKHLEK	542
Db	1121	DAVEHOKITAELEKTR	-----	IOL	-----	EYKSNQKIKEEYSNFORFKEOBOK	1168
Oy	543	LTKALTKMEDYSLRNIV	-	VEKEKTYKNLISKIE	-	NELEITVENIKKDEBQLEPKTK	599
Db	1169	RNSLIVESLNSKIKEL	EARLSOE	LSLNOYLNKISGNSVETNISTRK	-----	TS	1219
Oy	600	DENKPEKILLEVSDIV	YOVQVLN	MNKKIDELKQ	TLILKNELKHNHVPNSYKQENKQ	659	
Db	1220	YSDOP	-----	LDKEDILK	-----	KYUOLAFPEIRN	1247
Oy	660	EPPYLLIVKKEIDK	LKAFPMKVES	LNEEKKNIKTEBGS	DNSEPTGEITGOATTKQ	719	
Db	1248	-----	LENEI	-----	EEKKNLISLR	-----	FTETRL
Oy	720	QAGSALEGDSVOA	QAOBOKO	-----	AOPVAVPVEPKA	QOVPPAPVNNKTEVNSKL	772
Db	1270	ASSFPEDQIKKQMK	TKLTLID	MDPSITP	-----	DBILNEPLDNC	-DKESDINKLME
Oy	773	-DYLEKTYE	-	LNTSYCHKYILVSH	STNEKILKQYKITEEESK	SSCOPDLLENI	829
Db	1324	VDYLKROLDIETRA	HYDAE	MAISALHS	-----	KFRKI	-OGESSLSDDIYKILFEAS
Oy	830	-----	-----	ONNIPVMS	SMFDSLNN	-----	SLSOLMEIYKEMOYNLYKLDN
Db	1375	EERKSLSEDKL	KYMP	LDRKINLPV	-----	GDILKRDJSIK	-----
Oy	870	DKIKNLEAK	-	KVYSIVKTLSSSS	MOPLSLPQ	-DKPEV	SANDTSHSTNLNSLK
Db	1423	YKLOEITNES	NGKLSQ	LTDLROSK	SEKALISPO	DLROKDESTEROK	ELLSTRIKQK
Oy	925	-LEFNII	SLGKNNKIY	YELIGOK	SSENFEKILKSD	DTYNSFTNPK	SKRADDINSLD
Db	1483	QOFENCM	-----	-	DDLOGNLR	REHHA	KQAE
Oy	984	ESKERR	-----	LEEDINKL	KTYLOSP	LYNKKLELER	FDKKRYTGAKYKMOJK
Db	1526	ONQK	KEKILWREMER	NDSDMO	QETPL	-----	TELKRVQDYKILISDLAH
Oy	1034	-----	KLILKOLE	SKLNSLNNK	HYLONE	SVFFNKKKA	EIAETBNTLENTKILK
Db	1581	DRSOYT	DEINRKL	KEELNCS	LKAE	TNKL	-----
Oy	1087	HYKLYVYNG	ESSPLK	TEESIO	REDN	YASL	-----
Db	1637	HYTYVVEK	LNNEKDAIS	LAKELYO	-----	KTEALNTE	CSLKKIVSLTRIKIOE
Oy	1139	EKKRLSY	SSGLHHL	JAELEK	EVKNNKYGNSP	SENNMTDV	NNALYESYKFLP
Db	1694	KTDAL	IO	SNAL	-----	SSSTQK	NKRITETKI
Oy	1199	VSEGSOT	LEQSP	KRAKASHV	GAESNT	TTTSON	-VDEVDV
Db	1725	-----	TL	-----	QLOEONS	RNEELVTK	QASONGYKQKFD
Oy	1258	QVMTGE	AVTPSV	IDNII	LSKLENEY	EVLVYK	LAGVRS
Db	1767	E	-----	KOTLOK	INTDLOL	-----	QLNK
Oy	1318	RPNK	RENPKV	LES	DL	PKYKDLT	SSNYVVKDPYKFLNKE
Db	1801	KIHELE	NMVVS	-	LE	TDL	-----
Oy	1377	NFADV	IGYKILISE	KKSD	LS	IKKRYNDK	GENSKYLP
Db	1845	ELAK	NRSE	FEAR	LKEAD	ISPLEK	YISOOE
Oy	1437	IHEAK	VL	-	NYTYE	KS	NVEVKI

Db 1905 SRSESTMIGSKNIDSNNAQSKI 1926

RESULT 55

T18440

hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)

C|Species: Plasmodium falciparum

C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C|Accession: T18440

R|Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A|Reference number: Z18935

A|Accession: T18440

A|Status: preliminary; translated from GB/EMBL/DBJ

A|Molecule type: DNA

A|Residues: 1-4550 <LAW>

A|Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB1121.1

C|Genetics:

A|Map position: 3

A|Note: C0425w

Query Match	5.0%	Score 420.5	DB 2	Length 4550
Best Local Similarity	19.1%	Pred. No. 1.8e-05		
Matches 389	Conservative 288	Mismatches 679	Indels 685	Gaps 89
QY	72	SVASGSGGVSASGVSASGVSASGVSASGSGSRRTRNSDSDSDAKSYADLAKR	131	
DB	1683	SCSSSFWGSLYNTSDNISSEISNNLSSTSSDKSEH-YCSSISDHEIENDVYITRTYR	1741	
QY	133	VBNYLITLTKELKYPOLFDLTN-----HMLTLCDN-----	162	
DB	1742	INBRMKDGGKKNKSGKRDMSNKKGDKSCAKKKNICDNKNKIYDNKKLIYDNKKSSTDPAYE	1801	
QY	163	GFKYIDGKEEINELLNLYKTFEYEDLLRAKLVANCVAQOIPFNKRI-----RANEDVL	217	
DB	1802	NIDHLEI-KEEIDNLTNNIGLTYTNLKKNNNN---NVYMK---TILTKLCGTHSNEDH-	1853	
QY	218	KKLVEGRKPLDINIDNVGKMEDYIKKKKTITENI-----ELI	256	
DB	1854	-----KAVNNIKLTNT-----LNKKYNNINICIFHYHNSFLHSFYSEDEYIELI	1897	
QY	257	--EESKKTIDKKK-----	267	
DB	1896	KREERKEDIKKIRKGYKQHOEDRDNNNNNDNNNDNNNDNNNDNNNDNNNDNNNDNN	1957	
QY	268	-----NATKEEKKK-----LYQAOYDLSYNNKOLEEAHNLISYLEKRIDTLC	310	
DB	1956	NNNDNNNGEODLINEDKKRRKINNLSHTEHTHHNSDHKKRKKKKNKNEQOINDTJK	2017	
QY	311	KNENIKEL-----LDKINELKNPPRANGTPTNLL-----DNKKITIEHEKEIK	355	
DB	2018	KETHKNEIDHKQKDTNCCNGIEEKKDQYSKNNTIISKYHFHEFRKNNNIKKDEKKEXY	2077	
QY	356	ELAKTIKFNIDSLFDPLELEYLLEKKNKNIDISAKVEKESETEPEYNGTYPYSTYND	415	
DB	2078	ERNDKMS-ECMDLONKRDNDNNNNNNNDNDNNNNNDNDNNNNNDNDNNNNNDHNNNN	2136	
QY	416	INNALNELNSFEDLINPFEYTKEPSKNITYDERRKKFLNEIKER-IKIEKKKIESDQKSY	474	
DB	2137	NDDDNNNNNNN-----NDDDNKNKDPt---LQETHSYIDEINQKEINIFSKYIDKILENV	2187	
QY	475	EDRSKSLNDITREYKELKLENIYD---SKFNNNIDLTNEFKMKGRYSKYVEKLTHHTF	530	
DB	2188	EFKCKIMKNVNVCGIYLLTROENGLNICORVNNNI-----IKNMKQKEFKYKKRQKNKN--	2241	
QY	531	ASYENSKHLEKLTALKYEMEDYSL-RNIVYKELKLYYKNSLKISENETJLVENIKDE	589	
DB	2242	--HKKKKNK-----KYNIDINIFNNVPVIOKRN-----INTMININNNN	2281	
QY	590	EQLFEKKITTKDENKPEDEKILEVSDI-VKVOQVVLIM-----NKIDELKKT	634	
DB	2282	NNNSNNSNNHNNNNNSNH--INLSDLKKKINIKKITYTLFYRLPLFLGDKYINDIDDVKSI	2339	


```

QY 635 QLI-----LKNVELKLN-----HVPNSY---KOEKNQEP 661
Db 2340 YMIKNFNOYNDVNTFSNNILYIDLLKKYNTNHNILKKNMIQSPKVICKEEQSDP 2399
QY 662 YLLVLKKEIDKLVFMPKVESL-----INEKNIKTEGSD----- 699
Db 2400 AKNTYVFNHNAALTSHPVDTNTNLLSDVNIETNEIOEHINNIRDHNNVINNQOME 2459
QY 700 --NSEPTEGEITGOATTKPCOQ-----AGSALEGSVQAQAOEOKAOPRVNPEYA 751
Db 2460 PKNNNTNEEKIILESTTORSEQMEDREDEGDSGETRLVDVETKVVDVETKVVDVET 2519
QY 752 K-----AQPPTPAVNNKTEV----- 769
Db 2520 KVVVDVETKVVVDVETKVVVDVETKVVVDVETKVVVDVETKVVVDVETKVVVDVET 2579
QY 770 -----SKLDYLEKL-----YEFINTSY-----ICHKYLIVSHS 797
Db 2580 VDHKMYVDVHKMYVDVHKMYVDVHKMYVDVHKMYVDVHKMYVDVHKMYVDVHKMYVD 2639
QY 798 TMN-----EKILQYKITEE-----ESKLSSCDPLDLFNINNNIPMY 837
Db 2640 VVDVQITDAPNRKPIIDDCMKIKKKKKSETLVQEMEL-SCYP-----NKNVNIIF 2691
QY 838 SMEDSLNLSQLFMEIYEMVCNLKDKDKT--KNLLEAKKVSTVSKTSSSSMQ 895
Db 2692 SLFNGNNGN---CYLLIYEESE-----KDNKKKSGKNKALKTKTDKNNDIYEFKDL 2739
QY 896 PLSTTPDKPEVSANDTSHSTNIN-NSLKL-----FENILSLCKNNIYBELIGKSS 949
Db 2740 KI-----HEININEGSYKILKDIYFSMLVNMNMNMKNGIOYLNMRHR 2782
QY 950 NFEYKILKSDTFEYNESEFTNEVSKADINSLNDESKRKLEEDI-----NKLKKT 1000
Db 2783 SYNKKVSGKGS-----SNINNNINNNINNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2836
QY 1001 LQUSEDLNKKYKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 1060
Db 2837 EELNYIKFNE-----TILNK-----TILNN-----LKKKFG 2862
QY 1061 VEFNKKKEAEIETENTLENTLILKHYKGLVYKYNCGESSPLKT-----LSESIOTED 1116
Db 2863 C-LNITSAAH-KLVNILE-----LSYYDIYF-----LKHKKIPEFSKGVKIIDHS 2906
QY 1117 ASLENFK--VLKLEGLKDLNLEKRLKLSYSSGLHHL-----IAELK 1158
Db 2907 NNLTNFOENMNDSEYKMTKEFNKFLKFKKFRYLSLKLKILMSHCAGGLKTHNKKIGLK 2966
QY 1159 EVIKNNKVTGSPSENN-----DVNNALESYKFLPEGTDVATVVSSESSD 1205
Db 2967 GNSTHNNNNRNNYYNNTLNRGLASNNKKNNDMENIIMTYKKNNIINKETLNIIVTNGGL 3026
QY 1206 TL-----EOSOPKPKPASTHVG--AESNTIYT-----SON 1232
Db 3027 CLNNEGHOVARETEPLDINIRKVEKNNNTSHGILNMKNYNNMNMKNYNTSVHGIRDKN 3086
QY 1233 VDDEY-----DDVIIVIFGESEEDYDGLGVVTEAVTPSYIDNLSKIENEYEVLY 1285
Db 3087 CNAQLYDNTLNFNMKKKPKSGE--ENCVEDING--IDSKILTSDDKNNLNKNN--MKY 3141
QY 1286 LKPLA-GVYNSL--KOLEN--NVMTFNVNVK-----DIANSRPNKREN 1324
Db 3142 KNTLSKGVKTKLYCKKGGKKNKGNKYNVLIKTASLSSPLSPSYMNTINLNNNNNISHN 3201
QY 1325 FKNV-----LESOLIPYKOLTSNYYVVDYV--KELNKKERDKL--SSYNTIKSIDTDINFA 1379
Db 3202 NNNLILILMEYLNKNDINDINIKELNLKKNYNNFENYNNYDIKSTFVDHKIV 3261
QY 1380 NDVLGY-----YKILSEKYSDDLDSIK-----KYINDKOGENKYPFLN-----NIE 1422
Db 3262 PSVSSFSVFDLYKLFEPFDHK-ELDKINNEDOKMKYINDVLEFNKALNFRNFIMKKKHID 3320
QY 1423 TLKYTVNDR-----IDLF-----VHLE----- 1440

```

```

Db 3321 YVKKSEFNKNTIDIANSYLIDIIHLLSEKNQKLEFNKINNRRPTQEIOTIITHNNKNTI 3380
QY 1441 -----AKVLNTEKSNVEVKIKELNLYKTTIDOKLADFKNNNFVGIADLSTDY 1489
Db 3381 VYDNTKTEQDYCKILITLTDQON--NINHTNHTNDVKID-KKKKKKSVKNNVHMD 3436
QY 1490 NH-----NNLTRELSTGVWFENLAKTVLSNLDGNILOGLNLSIQHCVKKQCPON 1540
Db 3437 DHVDIKSIHVNN--TNKMNNTNKNMNTNKNMNTNKNM--NTNKNMNTIOMNNIKINSND 3493
QY 1541 SGCERHLDER--EECKC--LLNTKQEGDKCEVNPPLCNENNGCCADACTEEDSGSN 1595
Db 3494 ATPYIHTTQPEYESINCYAGLVRTCTSDKINTDYNVKNRENVSKNDLDVDDKDKTRN 3553
QY 1596 G 1596
Db 3554 G 3554

```

```

RESULT 56
A43854
major merozoite surface protein gp195 - malaria parasite (Plasmodium falciparum) (fra
C:Species: Plasmodium falciparum
C:Date: 10-Mar-1993 #sequence, revision 18-Nov-1994 #text, change 09-Jun-2000
C:Accession: A43854
R:Hui, G.S.; Hashimoto, A.; Chang, S.-P.
Infect. Immun. 60, 1422-1433, 1992
A:Title: Roles of conserved and allelic regions of the major merozoite surface protei
A:Reference number: A43854; MUID:92192814; PMID:1548068
A:Accession: A43854
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-89 <HUI>
A:Experimental source: FVO isolate
A>Note: sequence inconsistent with the nucleotide translation
C:Superfamily: major merozoite surface antigen

```

```

Query Match 5.0%; Score 417; DB 2; Length 89;
Best local similarity 97.8%; Pred. No. 2, 7e-07;
Matches 87; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 23 ESYQELVKKLEALDAVLTGYSLFQEKEMVLNEGTSHTAVTTPGSKGSAVSGSGSV 82
Db 1 ESYQELVKKLEALDAVLTGYSLFQEKEMVLNEGTSHTAVTTPGSKGSAVSGSGSV 60
QY 83 ASGGSVAAGSGSAVSGSGSAVSGSGNSRRT 111
Db 61 ASGGSVAAGSGSAVSGSGSAVSGSGNSRRT 89

```

```

RESULT 57
B71616
phosphatase (acid phosphatase family) PFB0380C - malaria parasite (Plasmodium falcipa
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence, revision 13-Nov-1998 #text, change 21-Jul-2000
C:Accession: B71616
R:Garther, M.J.; Tetteil, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: B71616
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2010 <GAR>
A:Cross-references: GB:AE001391; GB:AE001362; MID:93845168; PIDN:AACT71865.1; PID:9384
C:Experimental source: clone 3D7
A:Gene: PFB0380C

```

Query Match 4.9%; Score 414; DB 2; Length 2010;

Best Local Similarity 20.0%; Pred. No. 1.1e-05;
Matches 347; Conservative 283; Mismatches 588; Indels 516; Gaps 87;

```

Oy 2 KRIIFLCSLFEIIMQCTHESYOELVAKLEDAVLIG--SIFQEKVLANEGTSG 59
Db 635 KILF--CDELYIFN-----KILKRLQSLDMYKINHEVKGFGDKVLN----- 667
Oy 60 TAVTSTPGSKGSVASGSGSVASGSGSVASGSGSVASGSGSGNRRTNPSNSD 119
Db 668 -----NSYKCYDKNDYGSFSPYNKS 689
Oy 120 SDAKSYADLKHRVRYLLTIKELKYPQLFDLTNHLTLCDNIHG-----EKYLLDYE 172
Db 690 NDYKSH-----YIKKKMKNVKSVQCSNESIILEKQENEEKKKKKKKMMENF 738
Oy 173 EYNELLYKLNFEYDRLAKLNDVCAVNDYQIPNLIKIRANEDLVKLLVFG--YRKL 228
Db 739 NNNNLTMYNINFEFDLI--INERGNFQFYNNIKKKRQKNEKGLBEMVYNIIFOLMKYI 795
Oy 229 DNKDNVGMEDYIKKKKTIEN-----INELIESKKTIDKNKATKEEK 275
Db 796 LNEFSKFEFLFKLNKNVNDINTPSTINIKYIN--MYYHRKDCFEKQIHSKHHM 854
Oy 276 KRLYAQYDLSTYKQLEBAHNLISYLEKRIDTLKKKENIKELDKINEIK--NPPANG 334
Db 855 KRIH-----LRDKFIEYEKEN-----EIDNCNNINMDKKKEIN 889
Oy 335 NPPNLLDKNNKIEHEKEIKELAKTIKFNIDSLPTDPLELEYLREKKKNIDIS--AKYE 393
Db 890 NNYNNIDNNIIEIDMSNNFIETYYIIFYLLNMDTYIOFLYYL--KNTYILFSVAKVA 947
Oy 394 TKES-----TEPNPYNGVTPYPLSYN--DINNALNELNSFGD--LIMPDTKEPSKNIT 445
Db 948 EENSILMLTKLRKNHYIKLRNHHIINSVYKILN--NYKDEIFLYDYITK-----WT 999
Oy 446 D---NEKKFIEIEKIKIEKK-----IESDKSYEDRS-----KSLDITKEXE 489
Db 1000 ECKMATTDLIYNDVKNFTIDLENDIDIPITNDKEEYHVNNSIISVLKKNHNSVYKLLK 1059
Oy 490 KLLNEIY--DSKFN-----NNIDLTNEKMMGKRYSVKELTTHNHFASVENSKNHL 540
Db 1060 KIKNSIILKDLKLCNFTINKYIHTNTYDK--HNKIYDQIKIMWYH---PNNKKKNV 1114
Oy 541 EKLTRALKMYEDYSLR--NIVEKELKYK-----NLISKIENIETIYE--NIKKD 588
Db 1115 KIKKFIISYDAYIYHGVALNLNFNAYEKLQHPRSSIDILIKKEGQNNYIINGEIKKY 1174
Oy 589 EEO-----LEKKITKDEKNPRDEKILEVSDIVKVOQVYLLMMKIDELKKTOLI 637
Db 1175 EEONNFIIRPNINISGKMLSCHNKTNSNTLOGND-----REANILDADEBRLKRNKN 1230
Oy 638 LKNVLEKHHIYH--PNSYQENKQEPYLLIVLKKELDKLVMPKVESLINEEK--KNIK 693
Db 1231 QNRRAVYQNMNTNNNSKKRRNQTEYY-----KKEKKKKRYDQKNOQ--TNEQCAQKNO 1284
Oy 694 TEGQSDNSEPTEGEITGATTKPGQAGSALGDSVQAQAOBKQAOBPVPVPEAKA 753
Db 1285 TNEQ-----KNDQTNQTDNDQTDQ----- 1304
Oy 754 QVTPPAPVNNKTENVSKLDYLEKYLEFNTSY--IC-----HKYI-----LVS 795
Db 1305 -----TNDQTDNDQTN--DKIKRFYKNITCYTKLCKMEYSNKKYLSWLCSGMSLDIVY 1355
Oy 796 HSTAMEKILKQYKITEESKSLSSCDPLDLFNIONNIPVMTSMFDSLNNNIS--QLFMEI 854
Db 1356 NFIIIVRLYEKY--NKNENKTKTC-----PIRIILYLTGHOSLISFQSCVGI 1400
Oy 855 YEKEM-----VCNLYTKLNDKRIKMLLEAKKVSYSVTLLSSSQPISLTQPDOP 905
Db 1401 RKDKDKIPIPFASFISLELHIK--KKIKIKMLSKLCVNSNNEKSYCSNKNYNI--MKEEKK 1458
Oy 906 EVSANDDTSHSTNL-----NNSLKLFEINILSGKKNKIYQELIGQSSSENFYEKIL 956

```

```

Db 1459 HASSRSRVHNOQDRTDVLSEIFYHNNNTANIF-----CCKDCVWK--VRETENEKKEKOR 1511
Oy 957 KOSDFEYNSFPNFVSKADDINSLSNDESKRKLEEDIKKIKKTLQSLFDLYN----- 1009
Db 1512 KNK--KFMENEENAVIK--DDEKNITNLIK--NINENID--KKSININICITNDIPTNVA 1565
Oy 1010 --KYKLELRELPDKKRTVQKQKQKQIKKTLKBOLESKLSLNSNPNHVLONFPSVFNK-- 1065
Db 1566 NKKEYSLPRKCLNK-----INDPK--NLYLLOCKN-----NNIDLIQLYDCLNNNV 1612
Oy 1066 ---KKEAEIAPENTLENTKILIKH-----YKGLVYYNGESSPLKTLSESIQTEDNYA 1117
Db 1613 THIKKNMQKECK-----KHGRNFEYGFVFTPNNSVPLKRLKLLKKNK-- 1659
Oy 1118 SLENFVLSKLBCKLKDNLNLEKKKLSYSSGHLIAELAEVIRKKNKNTGNSPSPNMD 1177
Db 1660 -----MGKKDKDEEDN--NYHNDKNY--SDNLEYNHDTNN--NNNNNNNNNNNNNN 1709
Oy 1178 VNNALSEYKFFLEPGTAVTVVSESGSDTLEQSQPKRPASTHVGAESNTITTSQNVDEY 1237
Db 1710 NNICLKNNKNNT-----MHEDIANKRESLAKKKKK-----KKNCIQKNNNICERK 1756
Oy 1238 DDYI-----IYFGESEEDYDLAGOVVTEAATPSPVIDNLSKTIENEYVLYLK 1287
Db 1757 KSNIHNNSSKYIFNTVRF--KMKDIKAKINTKCD--ENSISCIINNRE----- 1802
Oy 1288 PLAGVYRSILKOLENVNVPFNVVKD--ILNSRPNKR-----ENFKVYVESDLIPYDLTS 1341
Db 1803 -----KRNIFKLNINILFNNSNNDKYNAYIYNSNTVYTKRYKIKNKDV-----H 1850
Oy 1342 SNYVADPYKFLNKRDRKFLSSYNIKDSIDTIDIFANDVLYGYKILSEKYSDDLDSIK 1401
Db 1851 INNILHTYK--QKKKKSTIIS-----DNNNNNNNAEDDISRKL--KFRDIGMFK 1901
Oy 1402 -KYINDQGENEYUPLPLNNIETLYTVNDKI--DLPYIHEAKVLANITYEKSANVEKI 1457
Db 1902 QKYINDH-----NNINSYDNNINNGLTNEHKVNLHNECK----- 1935
Oy 1458 KELNYLKTQDKLADPKKNNNFVGIADLSTDYHNNMLTFGLSTGWFENLAKTVLSNLL 1517
Db 1936 -----NKNQITG--YSIKYKN-----YSEKSCSDVITSLK 1966
Oy 1518 DGNLQGLMNSHQCVKQKOPNSGCFRLDEREC--KCLLNYKQSGDKVEMP 1570
Db 1967 DKKIKRR-----KKKIQKKN--YENENIVCLDCLISVLKMKMLRIYGNP 2007

```

RESULT 58

A56539

gi21411.1 human
N:Alternate names: macrogolin

C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence-revision 26-Jan-1996 #text-change 10-Dec-1999

C:Accession: A56539; #37536
C:Seelink, H. P.; Schanz, P.; Schroeder, H.; Wiemann, C.; Griffiths, G.; Renz, M.

Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein

A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539

A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>

A:Cross-references: EMBL:W75304; NID:g405714; PIDN:CA53052.1; PID:g405715

C:Genetics:
A:Gene: GDB:GOLGB1; GCP: GCP371

A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TM>

Query Match

Best Local Similarity 4.9%; Score 412; DB 1; Length 3259;
Matches 314; Conservative 301; Mismatches 592; Indels 376; Gaps 72;

[illegible][illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 31, 2003, 07:21:31 ; Search time 26 Seconds
(without alignments)
1854.776 Million cell updates/sec

Title: US-09-269-874A-3
Perfect score: 8424
Sequence: 1 MKRIIFLCSPFLFIINTQCV.....SNFLGISFLILMLITVSFI 1639

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 180 summaries

Database :
1: /cgn2_patents_AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfilltest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1824.5	21.7	377	4	US-08-195-705-5
2	1763.5	20.9	375	4	US-08-195-705-3
3	1289.5	15.3	394	4	US-08-195-705-2
4	1254.5	14.9	394	4	US-08-195-705-4
5	615	7.3	115	1	US-08-152-922A-7
6	448	5.3	10182	4	US-09-134-001C-3159
7	387	4.6	3696	4	US-09-134-001C-5080
8	356.5	4.2	3248	5	US-08-353-700-1
9	356.5	4.2	3248	5	PCT-US95-16216-1
10	353.5	4.2	2285	4	US-09-308-375-2
11	351	4.2	1786	4	US-08-973-462-8
12	350.5	4.2	2482	1	US-08-328-354-6
13	343	4.1	1939	4	US-09-310-187A-1
14	338	4.0	1886	4	US-08-938-105-3
15	332	3.9	976	4	US-09-104-324B-4
16	314.5	3.7	1151	4	US-09-134-001C-3242
17	310	3.7	1211	4	US-09-134-001C-4820
18	308	3.7	1312	2	US-08-687-080-51
19	307	3.6	1312	2	US-08-592-126-148
20	303.5	3.6	1038	4	US-09-541-782-4
21	303.5	3.6	1038	4	US-09-723-820-4
22	302.5	3.6	3169	4	US-09-453-702B-257
23	301	3.6	53	1	US-08-290-919-3
24	298.5	3.5	2710	1	US-08-480-604A-6
25	298.5	3.5	2710	2	US-08-405-496A-6
26	298.5	3.5	2710	4	US-08-915-136-6
27	298.5	3.5	2710	4	US-08-957-310-6

28	298	3.5	3135	1	US-08-323-170B-2	Sequence 2, Appl1
29	298	3.5	3135	4	US-08-954-441-2	Sequence 2, Appl1
30	296	3.5	1588	5	PCT-US93-07261-11	Sequence 11, Appl1
31	296	3.5	1663	5	PCT-US93-07261-16	Sequence 16, Appl1
32	294	3.5	1388	4	US-09-572-191-2	Sequence 2, Appl1
33	294	3.5	1388	4	US-09-723-262-2	Sequence 2, Appl1
34	294	3.5	1388	4	US-09-723-219-2	Sequence 2, Appl1
35	293.5	3.5	1010	4	US-09-134-001C-5178	Sequence 5178, Ap
36	293.5	3.5	2391	2	US-08-446-855A-2	Sequence 2, Appl1
37	293.5	3.5	2391	4	US-09-150-741-2	Sequence 2, Appl1
38	292.5	3.5	3418	4	US-08-755-587-44	Sequence 44, Appl1
39	288.5	3.4	3418	2	US-08-603-753D-4	Sequence 4, Appl1
40	288.5	3.4	3418	4	US-09-099-753-4	Sequence 4, Appl1
41	288.5	3.4	3418	4	US-08-966-106-4	Sequence 4, Appl1
42	282	3.3	53	1	US-08-290-919-4	Sequence 4, Appl1
43	280	3.3	48	1	US-08-290-919-12	Sequence 12, Appl1
44	279.5	3.3	1589	3	US-09-356-952-4	Sequence 4, Appl1
45	279.5	3.3	2329	3	US-08-755-587-16	Sequence 16, Appl1
46	279.5	3.3	3418	2	US-08-639-501-2	Sequence 2, Appl1
47	279.5	3.3	3418	3	US-09-044-946-2	Sequence 2, Appl1
48	279.5	3.3	3418	3	US-09-044-908-2	Sequence 2, Appl1
49	278.5	3.3	2366	1	US-08-480-604A-10	Sequence 10, Appl1
50	278.5	3.3	2366	2	US-08-405-496A-10	Sequence 10, Appl1
51	278.5	3.3	2366	4	US-08-915-136-10	Sequence 10, Appl1
52	278.5	3.3	2366	4	US-08-957-310-10	Sequence 10, Appl1
53	278.5	3.3	3788	4	US-09-336-447A-76	Sequence 76, Appl1
54	275.5	3.3	1093	4	US-09-315-793-52	Sequence 52, Appl1
55	274	3.3	48	1	US-08-290-919-1	Sequence 1, Appl1
56	274	3.3	1354	3	US-08-685-871-2	Sequence 2, Appl1
57	272	3.2	1535	3	US-08-755-587-185	Sequence 185, Ap
58	271	3.2	48	1	US-08-290-919-2	Sequence 2, Appl1
59	271	3.2	1164	4	US-08-923-992A-10	Sequence 10, Appl1
60	271	3.2	2154	2	US-08-841-349-4	Sequence 4, Appl1
61	270	3.2	2052	2	US-08-790-912-2	Sequence 2, Appl1
62	269	3.2	1435	2	US-08-568-459A-4	Sequence 4, Appl1
63	269	3.2	1435	2	US-08-487-826B-4	Sequence 4, Appl1
64	269	3.2	1435	2	US-09-210-288-4	Sequence 4, Appl1
65	267.5	3.2	1388	2	US-08-685-576-4	Sequence 4, Appl1
66	266.5	3.2	1579	3	US-08-755-587-184	Sequence 184, Ap
67	266	3.2	1164	4	US-08-923-992A-2	Sequence 2, Appl1
68	265	3.1	1363	1	US-08-425-061-23	Sequence 23, Appl1
69	265	3.1	1363	2	US-08-825-886-23	Sequence 23, Appl1
70	265	3.1	1852	1	US-08-425-061-24	Sequence 24, Appl1
71	265	3.1	1852	2	US-08-825-886-24	Sequence 24, Appl1
72	265	3.1	1863	1	US-08-425-061-16	Sequence 16, Appl1
73	265	3.1	1863	1	US-08-480-784-2	Sequence 2, Appl1
74	265	3.1	1863	1	US-08-483-553-2	Sequence 2, Appl1
75	265	3.1	1863	1	US-08-487-002-2	Sequence 2, Appl1
76	265	3.1	1863	1	US-08-483-554B-2	Sequence 2, Appl1
77	265	3.1	1863	1	US-08-488-011B-2	Sequence 2, Appl1
78	265	3.1	1863	2	US-08-825-886-16	Sequence 16, Appl1
79	265	3.1	1863	4	US-08-850-727-2	Sequence 2, Appl1
80	265	3.1	1863	5	PCT-US95-10202-2	Sequence 2, Appl1
81	265	3.1	1863	5	PCT-US95-10203-2	Sequence 2, Appl1
82	265	3.1	1863	5	PCT-US95-10220-2	Sequence 2, Appl1
83	265	3.1	1964	2	US-08-790-912-3	Sequence 3, Appl1
84	264	3.1	1863	1	US-08-798-691-6	Sequence 4, Appl1
85	264	3.1	1863	3	US-08-825-487A-4	Sequence 4, Appl1
86	264	3.1	1863	4	US-09-074-476-6	Sequence 6, Appl1
87	263	3.1	1863	1	US-08-598-591-2	Sequence 2, Appl1
88	263	3.1	1863	1	US-08-798-691-2	Sequence 2, Appl1
89	263	3.1	1863	1	US-08-798-691-6	Sequence 6, Appl1
90	263	3.1	1863	3	US-08-825-487A-2	Sequence 2, Appl1
91	263	3.1	1863	3	US-08-825-487A-6	Sequence 6, Appl1
92	263	3.1	1863	4	US-09-074-476-2	Sequence 2, Appl1
93	263	3.1	1863	4	US-09-074-476-4	Sequence 4, Appl1
94	262.5	3.1	106	1	US-08-290-919-11	Sequence 11, Appl1
95	260	3.1	740	1	US-08-257-073-5	Sequence 5, Appl1
96	259.5	3.1	1388	2	US-08-685-576-1	Sequence 1, Appl1
97	257	3.0	1332	4	US-09-345-882-29	Sequence 29, Appl1
98	254	3.0	1863	2	US-08-603-753D-2	Sequence 2, Appl1
99	254	3.0	1863	4	US-09-099-753-2	Sequence 2, Appl1
100	254	3.0	1863	4	US-08-986-106-2	Sequence 2, Appl1

[illegible]

Db 3399 AIRKEFAIAEIOELADKK-----LNKFKO-----SQEAT-IEBKQAI 3435
 QY 1207 LEQOPPKPASTHVGASNTTTSQNDVDDVYIIVIFEESE-----EDVDDLGCV 1259
 Db 3436 NELBOALKSATNHHSOSNNESVSAALKESTLSIDSEIOAHKKLEAKAYIDGSD---- 3491
 QY 1260 VTGEAVPTSVINDILSKIEENEVLYLPLAGVYRSKLQLEN-----NWTFFVNVK-- 1312
 Db 3492 -----DKINDISSRATNEKQIFVSKLALINRHKQIDAEFTFVSVEIYRNFKEVE 3543
 QY 1313 -DILNSPKNKRNKNVLESLLIPYKDLTSSNYV---VKDPYKFLNKKKDKLSSYNY 1367
 Db 3544 ADKNLSTYRKAKASKETELEADHVKOMINANLSTAFVKONARTLLINEIVSNALSQLNK 3603
 QY 1368 IKSDIDIDINRANVLYGYKILSEKYSKSDPSIKKYYINDKQGENKVEYLPFLNIETLYKT 1427
 Db 3604 VTTKKEVD-ELVNETIEKLKSIQIREDKILSSQSSSTSMTEKSNOCYSSENNITKSLPEA 3662
 QY 1428 VN-DK 1431
 Db 3663 GNADK 3667

RESULT 8

US-08-353-700-1
 ; Sequence 1, Application US/08353700
 ; Patent No. 5599919
 ; GENERAL INFORMATION:
 ; APPLICANT: YEN, TIMOTHY J.
 ; APPLICANT: RATTNER, JEROME B.
 ; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 ; TITLE OF INVENTION: TRANSIENTLY EXPRESSED KINETOCHORE PROTEIN,
 ; TITLE OF INVENTION: AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DANN, DOFFMAN, HERRELL AND SKILLMAN
 ; STREET: 1601 MARKET STREET, SUITE 720
 ; CITY: PHILADELPHIA
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103-2307
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/353,700
 ; FILING DATE: 09-DEC-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: REED, JANET E.
 ; REGISTRATION NUMBER: 36,252
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 563-4100
 ; TELEFAX: (215) 563-4044
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3248 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: HUMAN
 ; US-08-353-700-1

Query Match 4.2%; Score 356.5; DB 1; Length 3248;
 Best Local Similarity 18.9%; Pred. NO. 1.2e-10;
 Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADLKHR--VNNYLLTTELKELYPOLFDLTNNMLTLCQNIHGFYKYLDS---YEENINELLY 179
 Db 709 FSDQKHQKEIENMCKLTSQLT-GQVEDL-EHKLQLLSN-----ELMDQDRYQDLHAHYE 761
 QY 180 KLNFEYDLRAKLANDVCAND-----YCOIPENLKITRANELDVLKLVFGYKRPDLNI 231
 Db 762 SLR--DLKSKDASLYVNEHQSLAFDQAPAMHHSFAN-----IIGEGSWPSE 810
 QY 232 KDNYGKMDYIKKKKTTEINENNELLESKTTIDKNKNAITREKKKLYQAOYDLSTYKQ 291
 Db 811 RSECRLEADQSPKNSAIIQNNVDSLEFLES-QKOMNSDLQCEBELVQINGEILENMK 869
 QY 292 LEEAH-NLISYLEKRIDRLKKN-----ENIKELDKINEIKNPAPNSGNTPTLL 341
 Db 870 AEOHQSVAETSGRIKSLQEDTSAHQVVAETLSALENKEKELO-----LL 916
 QY 342 DKNKIEHEKEIKEIAKTIFNIDS-----LFPDPLEYLYLRKNKNIDISAKVETKE 396
 Db 917 --NDKVEEQAEIOELKKNHLEDSLKEIQLLSETLSLE--KKEMSIISLN-KREIEE 971
 QY 397 STEPEYNGVYTPYSTYNDINNALNELNSFGDLLNPFYTKPEPSKNITDNERKKFINEI 456
 Db 972 LTOENG-----TLKEINASLNQEKM--NLIO-----KSEFANTID-EREKSISEL 1014
 QY 457 KEKIKIEK---KRIESDKSYEDRS--KSLNDITKYEKLEINEYDSKFNNNIDLTNF 509
 Db 1015 SDQYKQEKLIILQREELGNAVIEDLSQYKKAQENKNSLQEDLNDCTSLCENRKNLEQL 1074
 QY 510 EKMGGKRYSYVEKLTHTNFTASTYENSK--HLEKLTALK-YWEDYSLRNIVVEKELKY 566
 Db 1075 KEAPAKEHQEFLTKL---AAAEERNQNLMELETVQALRSEMTDQNNKSKSEAGIKQ 1130
 QY 567 YKNLSKIEENIEFTLYENIKKDEQLEPKKITKQE-----NKRDEKILEVSDI-- 614
 Db 1131 EIMTLKEQONKQKEVNDLQENDQMLKVMKTKHBCQNLSEPIRNSYKERESEKQCNF 1190
 QY 615 --KVQYQAVLNMKIDELKTKQILKLNVEIK-----HNHV----- 649
 Db 1191 KPQMDLEKVEKISLDSYNQALQVLEAMLNKELKIQESEKKECQHLQTLRGDLETSLN 1250
 QY 650 -----PNSYKQENKOEPIYLLYLKKEIDKLKVPKPVSL 684
 Db 1251 QDMOSQELISGLKDEIDAEERYISGPHLSTQNDNALQCSLOTYNNKLN-EKICEI 1309
 QY 685 INEKNKNTKTEGQSDNSPSTEGEITGOATTKPQOAG-----SALEGDSQV 731
 Db 1310 LQAEKVELVTELNDRSRCIT-----ATRKMAEYVKLNEKYLNDSDGLHGLVE 1362
 QY 732 ---AQAOEQKQAPVPVPPVPEAKAQVTPPAPVNNKTENVSKLDYLEKLYEFLNTSYI 787
 Db 1363 DIPGEFGEQENQHPVSL-----APLDEGNS-----YEHLTLS-- 1396
 QY 788 CHKTYLVSHTNNEKIL---KQYKTKREESKLSCDPLDLFNINQNIPIWYSFDSL- 843
 Db 1397 -DKEVQMHFAELOEKFLSLQSEHKILHDQHCQMS-----KMSLQTYVDSLK 1443
 QY 844 -NNSLSQFMELYEKEMVCNLYKLKDNDKIKNLLEAKVSTSVKTLSSSMQ----- 895
 Db 1444 AENLVLSNLNRFQGLVKEQQLGEBGLVPSLSSCCVPDSSSLSDQSSFTYRLLBQT 1503
 QY 896 -PLSLTPQDKPEVSAN-----DDTSHSTNLNLSLKFENILSLG 933
 Db 1504 GDMSLSLNLBEGAVSANGQSVDEFCSSLQTYVDSLAKAENLVSLNLRNFQGDVKEQMG 1563
 QY 934 KKNKIYQEL-----IQKSSNFYEKILK---DSGTFYV-ESTTNVYKSKADI- 978
 Db 1564 LEEGLVPSLSSCVDPSSSLSLGDSFYRALLBQTDGMSLSLNLGEGVANSANGQSVDEVF 1623
 QY 979 -NSLNDSESKRRK-----LEEDINKLTKTLQDSPLYKKYKIKLERL 1018
 Db 1624 CSSLOEENLTKRPPSAPAKAVEELESCEVYRSLERLEKEMSGQITMKKKEIOELQL 1683
 QY 1019 FD-----KKTVGYKKQIKKLTLLKLEOLSKLNS-----LNNPKHYLVN 1058


```

Db 1684 LSSERQELDLCKROYLSNEMOQKLTSTVLEMSKLAERKQETQSLSELAVARLOLG 1743
QY 1059 FSIV-----FFKKKEAELEAENTLENTKILKKH-----1087
Db 1744 LLSRLSLGLIDTEDAIOGRNESCDISK-EHTSETTERPKHVOICDKAQQDLNLDI 1802
QY 1088 ----YKGLVK-----YNGESSP-----LKTLS-----1107
Db 1803 EKITEGAKPKPGCEGSESPDPTNPEPREDKQSSSECSISLSTFGPNALVPMDLGNO 1862
QY 1108 ---ESTQEDNATSLNFKVLSKLEG---KLDNLNLEKKKLSYSSGLH---HIAEL 1157
Db 1863 EDIHNLQLVKETSNEHLRLHIEDRDKRVESILN---EMKELDSKLHLEVOQMTKI 1918
QY 1158 KEVKKKNTGNSPSNNTDVNNALFSSY-----1185
Db 1919 EACIELEKTVGLKKE-NSDLSEKLEYFSCDHOELLQRYETSEGLNSDLEMAHAKSSRD 1977
QY 1186 -----KKELPGSTVAVATVSSGSDTLLEQSOPKPKPASTHVAESNTITTSQ- 1231
Db 1978 IGDNAKAVDSMKERFLDVNELSRISSEKASIEHE-----ALYLEADLEVOQTEKI 2029
QY 1232 --NWDEVDVYIIVPIFGESSEDDYDLGOVVT-----GEAVTPS-----VIDNLSKIE 1278
Db 2030 CLEKNENKQKVIYCL-----EDEL-----SVATSERNQLRGELDTMSKTTALDQLEKKK 2081
QY 1279 -----NEVELYLYKPLAGVYRSLKQLENNVMTFNVVKDILNSRFNKNRENFNVIE 1330
Db 2082 EKTQELSHSQSECHICQVAAEAVKRETEL---LQTLSDVSELKDKTHLOEKLOS-LE 2137
QY 1331 SOLIPYKDLTSSNYVVKDPKFLNKKR-----DKFLSSYNYIKOSIDTIDINPAND 1381
Db 2138 KD---SQALSTLKCELENOIAOLNKEKELVKESESLQARLSSEDEYKENSALBAALV 2194
QY 1382 VGGYKI-LS-----EKYKSDLS-----IKKYNKQGENEYKLEPFLNN 1420
Db 2195 EKGEFRLRLSTQOEYHQLRGIEKLRVRIEADKKOLHIAELKREKERNDSLKQVEN 2254
QY 1421 IETLTKTVNDKIDLEFVILHLEAKVLTYYKSNVE-----VKIKEL-----NYLKT 1465
Db 2255 LRELOLMSSENGELVILDAENSKAEVETLKTQIEEMARSLKIFEDLVTLRSEKENVLQO 2314
QY 1466 IODKLADFKNNNNFVIGIADLSTDYNNHNLITKFLSTGMVFNELAKTVLSNLDGNLQGM 1525
Db 2315 IOEKQQLSE-----LDKILLS-----FKSLLEKEQOAEI 2344
QY 1526 NTSOHQCVAKKQCPNSGCFRHLDERECKCLNYKQEGDKVENP-NPTCNENN 1578
Db 2345 QIKESKTIKAVEMLQN-----QIKELNEAVALLCGDOEIMKATEQSDPPIEBEH 2393

```

RESULT 9

```

PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Ratner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 4.2%; Score 356.5; DB 5; Length 3248;
Best Local Similarity 18.9%; Pred. No. 1.2e-10;
Matches 350; Conservative 304; Mismatches 631; Indels 569; Gaps 79;

QY 125 YADLKHR--VRNLLTIKELKYPOLFDTLNNHMLTCLDNIHGFIYLDG---YEEINELLY 179
Db 709 FSDQKHQKEIENMCLTSTLT-GQVEDL-EHKLQLSN-----EIMDKRCIYDLAAEVE 761
QY 180 KLNFEYDLRAKLDVNCAND-----YQAIPENLKIRANELDVLKLVFGYRKPLDNI 231
Db 762 SLR---DLLKSKDASLVNEDHQRSLAFDQGPAMHSPAN-----IIGQSGMPSE 810
QY 232 KDNVGMKEDIYKKNKTTIENINELIESKTTIDKNNNAKREKKKLYQAQYDLSTYNNQ 291
Db 811 RESCRLEADQSPKNSAIIQNRVDLSLEFLES-OKQNSDLQKCEBLVQIKGEIENLMK 869
QY 292 LEEAH-NLISVLEKRIQDTLTKN-----ENIKELDKINIKKPNPANGTNPFLI 341
Db 870 AEGMHQSPVAETSQRIKSLQEDTSAHQNVVAETLSALENKEKELQ-----LL 916
QY 342 DKKKKIEHEKEIKELAKTIKENIDS-----LETDPLEEYVLRKKNKIDISAKVETRE 396
Db 917 --NDKVETQAEIQELFKKSHLLDSLKELOLSTLSTLE--KKEWSSIISLN-KREIEB 971
QY 397 STEPNFEPNGVTPPLSYNDINNALNELNSFGDLINFPDITKPSKNITVDNEKKKINEL 456
Db 972 LQENG-----TLKEINASLNOEKM--NLIO-----KSESFANYID-EREKSISEL 1014
QY 457 KERKIEK---KRIEDSKSYEDRS---KSLNDITREYKELLENLEYDSKFNNDILJTFN 509
Db 1015 SDQYKQEKLLIQRCEFTGNAYEDLSQKKAQEKSKLECLINETSILCENKKNLEQD 1074
QY 510 EKKMGKRYSYKVEKLTHTHTFASYENSK--HNEKLTALK-YMEDYSLRNIVYERKELY 566
Db 1075 KEAFKEHOEFLTKL---AFAERNQNLMELETFVQALRSBMTDNQNNKSEAGLQO 1130
QY 567 YKNLISKIENETIETVENIKKDEQLFEKKITMDE-----NKPDEKILEVSDI-- 614
Db 1131 EIMTLKEBQNKQKQEVNDLQENQOLMKVYKTHQCONLESEPIRSVYKRESERNQCNF 1190
QY 615 ---VKQVOVKVLLMKNKIDELKKTOLIKLVNELK-----NHIV-----649
Db 1191 KQPMDEVAKEISIDSYNAQVLEAMLRNKKELKLOESEKEKECTQHELOITRGDLETSM 1250
QY 650 -----PNSYKQENKQEPYLLIVLKKEDIKLVKFMFVRESL 684
Db 1251 QDMQSOEISGLKDCETDAEKEYISGPHELSTQNDVAHLQCSIQTTMKNKLE-LEKICEI 1309
QY 685 INEKKNITEGSDNSSESTEGEITGOATTKPGQAG-----SALBEDSVQ 731

```


Dd	1310	LQAEKVELYELANDSREBCT	-----	ATRKAAEEVGNLNEVKIILDDSGILGELVE	1362			
Qy	732	---AQAOEQKQAPVAVPEPAKAQVPTPAPVANKTKENYSKLDYLEKYEFLNITSYI	787					
Dd	1363	DIPGGEQEOQNEQHVPVL	-----	APLDESNS	-----	YEHITLTS	1396	
Qy	788	CHKIIIVYSHSTNMKIL	---	KQYKTKREBSKISCDPLDLFLNQNNI	PWYMSFDSL	843		
Dd	1397	-DKVQNHFAELQEFKFLSLQSEHKILHDHCOMSS	-----	-KMSLQOTVDSLK	1443			
Qy	844	-NNSLSQJPMFEYKEKVCWLYKLIKONDKIKNLLEAKKVTSTYKTLSSSSMQ	-----	895				
Dd	1444	AENLYVSTNLBNFOGDLVEMQGLEBGLVPLSSCVPDSSLSLGDGSSYRALLBQOT	1503					
Qy	896	-PLSLTPQDRPEVSAN	-----	-DPTSHSTMLNLSKLFLFNLSIG	933			
Dd	1504	GDMSILSLNLEGAVSANOCSDVEYFCSSIQOTYVDSJKAEULVSTNLBNPOGDLVEMQLG	1563					
Qy	934	KNKNIIOYEL	-----	-IGOKSENFEFKLIK	---	DSDTYTN	-ESTFTNFKSADDI	978
Dd	1564	LEEBGLVPLSSSCVDPDSSLSLGDSSFPYRALBQOTGDMSLSLNLEGAVSANOCSDVEYF	1623					
Qy	979	-NSLNDSEKRRK	-----	-LEEDINIKTKTLOLSFDLYNKYKIKLERL	1018			
Dd	1624	CSSLQEBELTRKETPSPAPAKGVEELBSLCEVYRQSLERKLEEMBSOGIMKNKEIOETLEOL	1683					
Qy	1019	FD	-----	-KKTKYKQMYQIKTKITLLKEOESKLNS	-----	-LNNPKHYLON	1058	
Dd	1664	LSERQELDCLRKROYLSENQMOOKLTSYLTENESKLALEKKOTQOLSLELEVARLOLOG	1743					
Qy	1059	FSV	-----	-FNKKKEAETAEENTLENTKILKKH	-----	1087		
Dd	1744	LDLSRSRLIDIEDAIOGNESCDSIK	-EHTSETTEPTPKHDVHQICDKQAQOOLNDI	1802				
Qy	1088	---YKGLV	---	YNGESSP	---	LKTLSE	---	1107
Dd	1803	EKITETAVNPTGECSSQSPDTNPEPGBEDTKQSSSEICISLSPSGNALVPMDFLQO	1862					
Qy	1108	---ESIQETENYASLENFVKYLSKLEG	---	KLKDNILKEKKIKSYSGHL	---	HLIAEL	---	1157
Dd	1863	EDHINLOLRKRENSNEMRLLIHYIEDRDKVESLNL	---	EKKELDSKLHLOEYQVLMKI	---	1918		
Qy	1158	KEVIKKNKYTGNSPSENNDVNNALESY	-----	1185				
Dd	1919	EACIELEKIYIGELKE	-NSDSLSEKLEYFSCDHQELLQVETSEGLNDEMNAOKSSRED	1977				
Qy	1186	---	KKFLPESTDVAITYVSSGSDTLEQOSOPKRPASTHYGASNTITTSQ	1231				
Dd	1978	IGDNVAVNDSMKERFLDENELSRISREKASIEH	---	-ALYLEADLEVVOETKL	2029			
Qy	1232	---ANVDEVDVITVPIFGSESEEDYDLOGVY	---	GEAVPSS	---	VIDNLSKIE	1278	
Dd	2030	CLEKNENKOKKAVYCL	---	EEBEL	---	SVTSSERNOLRGELIDTYSKTTALDOLSEMK	2061	
Qy	1279	---	NEEYVLUYKPLAGVYRSKLQDENNVMTFNNVYDIINSFRKNREKNKYLE	1330				
Dd	2082	EKQTELSHQSECHQICQVAEAVKRETEL	---	LQTLSSDYVELLKDTHIQEKIOS	-LE	2137		
Qy	1331	SDLIPYDLTSSNYVKDYPKFLNKEK	---	-DKELSSYNYIKOSIDTIDINFAND	1361			
Dd	2138	KD	---	SOALSLTCELENOIQAOLNKEKELLVKESELOARLSESDYKELTANSKALEALY	2194			
Qy	1382	VLGYYKI	-LS	---	EKKSLDLS	---	IKKIYIDKQGENEKLPLFLNL	1420
Dd	2195	EKGFEALRSLSTOEYHQLRGIEKTRVRIEADKKOLHIATKLERERENDSLKDYEN	2254					
Qy	1421	IETLYTKVNDKIDLEFVLIHEAKVLYNTYKESVNE	---	VKIEKL	---	NYLKT	1465	
Dd	2255	LERLOJNSEENQELVILDAENSKAEVETLKTQITEBANSIKTFEIDLVTLSKXENLTKQ	2314					
Qy	1466	IQDKLADFKKNNNFVGIAIDLSTDYNNHNLTLFSLTGAVENLAKTVLNLNLDGNLOGM	1525					

```

Db 2315 IQEKQGLSE-----LDKLLS-----FSLSEKQOAI 234
Qy 1526 NISQHCYVKQKQPNNSGCFRHLDERECKCLLYKQEGDKVEMP-NPTCENN 1578
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2345 QKESKTAIVAMLQN-----QKELNDAVALGDDQEIKNATFQSLDPREEH 2393

RESULT 10
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCF
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match          4.2%; Score 353.5; DB 4; Length 2285;
Best Local Similarity 20.1%; Pred. No. 12e-10;
Matches 311; Conservative 244; Mismatches 546; Indels 443; Gaps

Qy 119 DSDAKSVADLKHRYNRYLLTIKEL-VPDQEDLTNMLTLCDN-----HGFKYLDGEYBI 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 874 DSAKTFEDASKKEIK---SHDELQYKQIADYNDKGRPMWDLIADDDYKVAAD---KA 927

Qy 175 NELLYKLNFTYDLLRAKLND---VCANDYQIPEPNLKIRANELDYKKLYVGYRKPLDNI 231
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 928 KQSMKKAQSDIESGNNAKVQDSVLSIAAAYSIDISMTLXTSISDVNKL-----NL 978

Qy 232 KDNQGMKMDYIKKKKKTITENNELIEKSKTTIDKNNNAKREEKKLYQAOYDLSTYKQ 291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 979 KQDLDPEE---LEKFSSTIGKIQEKMQALDGGCEKAFDNKKDL---QSLLETYSK- 1029
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 292 LEEAHNLISVLEKRIIDLKKNENIKELLDINEIKNPSPANGMPNTLLDNKKRIEHE 351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1030 ---SDSSIDVFKMSFD---KMQKNIKQDCKSLSSVK---SEVGDLDGETLAEGNNAEDPG 1080
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 352 KEIKEIAKTIKFNIDSLFTDPLELEYLRKKNNDISAKVETKESTEPENYPCGVTPPL 411
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1081 KKLKEAL-----DANSVDIDIKAAI--KEMSDAMQF-DSVQDVL 1159

Qy 412 SYNDNLNALNLSFGGLINPFOYTKPEKSN-----IYTDNERKKFINIKIKIKIEKK 466
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1116 NGDIFNNPKDVAFLNDLLEKMAEGSISISANEANTLIQDKLELAQISLENGCVAINDE 1179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 467 IESDKSYSDERKSLNDITKEYEKELNLEIYDSKFNNNIDLTNPEKKMGKRYKYKEKLTN 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1176 VIKQRKVKLD---AYNDMVTYSKLMK---TEVNNAIKTLNADLT-----RIDSL- 1219
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 537 HNTFASYENSKHNLEKLTJKAUKYMEDYSLNIV-VEREKLY-----YKNLSIKIE 575
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1220 -----KKLKRERKRLDMGEAEISLDELYKSIINNVAADAKKELKLEKKMLDGGYSN--SQIE 1272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 576 ---NEIPELVLEIKADEQOLEFEKKITIKADENKPDEKILLEVSDIYKVQYOKVLLMNKIDELK 633
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1273 AMQSKSALESYISASE---EATSTOEMNK--QALVEAG-----TSLNMTWDOOK 1318
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 634 TQLLKINVELKHNHIVENSYKOE-----NKO-----EPYLIYLVKKEIDYK 675
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1319 A-----NEETKTSMTYVDKKEALKEKVAEIDKYNAKQVNDYKRYISOKRIDAKKEIKALQ 1373
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 676 VFMRKVESLLINEKKNIKTEGQSDN-----SEPTGEGLITGQATTGPGQO 720
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db 1374 ----OKKLMOEQAKLLKDOIKNSTGYOYVTTSSGGTSPSGSYSGKYSYINSA 1429
QY 721 ACS-----ALEGDSVOAOEOEKOAPVPVPEAKAOPVPBPAD-----VANKTEENVS 770
Db 1430 ASKYNDPPLIAAVIOEGSFGNAKASGV-----GAMGLMQLPALAKSLGIVNNAIDPYO 1484
QY 771 KLDYLEKLEYPLNTSYICRKYIILVSHSTNEKILKOYKITTKEESKLSGCDPLDLLENIO 830
Db 1485 NV-----MGOTKYLAQO--LEFGGVNEKALAAVYNAAGPQVNIKYGGIIPPEKTONYV 1534
QY 831 NNIPVWYSMFS-----LNNLSQLFMEIYEKENCVCNLYKIKLNDKXIKNLLEAK 880
Db 1535 KKIIMANYSLSLSATSSIASYTTNNSAFRYSKYGOEGSLRSSPHKGTD-----FAA 1587
QY 881 KYSTSVKLTSSSSMOPLS-----LTPOD-----KREVSAND-----911
Db 1588 KAGTAKLSQSGKVOIAGYSKTAGNWNVIKODGTAVAKYMHMLNTPSVAKAGSVKAGOTI 1647
QY 912 ----DTSHTNLLNSLKFENILSLCKNKNITYOELIGOKSSENFEYKILKSDPTFYNEF 967
Db 1648 GRVYSTGNSTGNHLQIEQNGKTIIDPER--YMOGIGTS-----ISDAQAEAEERQ 1696
QY 968 TNFVRSKAD-----DINSUND-----ESKRKLEEDINKLKTQLSFDLYNKY 1011
Db 1697 QGIQAOKSDLSLQGISSVNDIOILOYELVOSKLDPEDKRIGDDVRIAKDESMAKRY 1756
QY 1012 ----KLLERLPDKKKTGVRY--KMOIKLTLLEQESKLSLNNPKHVLNPSVFEK 1066
Db 1757 TSDSKFEFRYTSQOKKAVAEQAKIOQOKVMIQKEIKTN--KALNS-----1800
QY 1067 KKEALMETENTLENTKIILKHKGIVKYNGESSPLKTLSEESIQI--EDNVASLENFKY 1124
Db 1801 --AQRQLOEELKQAKLIDLSVODV-----RELQOLOVOSKVBETLKSIE--KS 1846
QY 1125 LSKLEKILKDNLNL-----EKKKLSYLSGHLHLAELKEVINKNNYTSNPSENNTOV 1178
Db 1847 SSKTQOKIKDVNKKISMTEDEDEKVKYISQOILLOQOQKEA---KKYI-----1892
QY 1179 NNALESYKFLPEGTVAIVVSESGDLEOSQPKRPASTHVGAE--SNTITTSQNVDEY 1237
Db 1893 --KOLEBQKKAQKFPDIOEQIT-----EEMQNMKQKQFMLEYNTKSKIKDIKYSL 1944
QY 1238 DQVIYPIGSESEDDYDLGOVVTGEAVTPSVYDNLISKENYEYVLYLKLPLAGVRSKL 1297
Db 1945 ABE-VSITIKEMEKMRDI--ELEAHQAKQDLDLDEL--DKTDE-----AKFOKELK 1992
QY 1298 KOLEN--NVMTFNVNVADILINSFKNREKFNKLVESLDLIPYKDLTSSSNVYKDKYKFLNK 1356
Db 1993 ERQDSIQIKLTDQINOYSLDDSEFGSK-----VKELTEQ-----LQKE 2030
QY 1357 KRDKFLSYNYIKDSIDTDINFANDVLYGYKILSEKYSDL--DSIKKYINDKQGENEKYL 1415
Db 2031 Q-----LDLDDFLKDR-----ESNKRKEALQDLEK--DEESSINNKYD 2066
QY 1416 PFLNNTIETIKKYVNDKIDLFVHLEAKVNLNTYEKSNEVEKIKELNYLKTIOKLADEFK 1475
Db 2067 NLVND--ERAFKFKLEDKI-----MNGKITDIK 2092
QY 1476 N--NNEVGIALDSTYNNHNLTKFSLSTGWFENILAKTVLSNLD 1518
Db 2093 QLNLF-----SKFINTNM--ESIGKSISNNLID 2118

```

```

RESULT 11
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRILLHE, PIERRE
; APPLICANT: DUBBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B

```

```

; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

```

Query Match 4.2%; Score 351; DB 4; Length 1786;
Best local similarity 20.0%; Pred. No. 1.2e-10;
Matches 269; Conservative 243; Mismatches 437; Indels 398; Gaps 68;

```

```

QY 229 DNIRKDVGMEDYIKRKNKTIENINELIES--KTIIDKNNAKTEKKKLYQAQYDLST 287
Db 757 ESVEENV--ESVAENVE--ESVAENVEESVAENVEESVAPEVEIIVAPSVESVAPSV 811
QY 288 YNKOLEE-----AHNLISYLERIDPLKKNENIKE--LLDKINIKNPPRANSQNTFTIL 341
Db 812 EESVAENVATNLSNDNLNLGIGIET---EETKDSLNIIEVKE-----NVVTTIL 860
QY 342 DKKKKIIEHEKE-----IKEIAKTIKENIDSLFTPDLELEYLREKKNIDISAKVETKES 397
Db 861 E--NVEETIAESVTFSNLEIEIQEN--TITNDTIEEK--LELHNEN--LSAALENTOS 912
QY 398 TEPNEYPNGVYTYPLSINDINNALNELNS--FGDLINPDYTKEPSKNYIYDNEKKAFINE 455
Db 913 EEEKK-----EVDVIEEVEKEEVAATTLIETVEQAEKKSANT-----ITE 951
QY 456 IKKIKIEKKIIESDKSYSDRSK-----SLNDITKYEKILNLEIYOSKFNNDITLTFEK 511
Db 952 IFE--NLEENAVSNEVAVENLEKLNETHVNTYLDKVEEVEIYSGESLENEMDKAFSE 1009
QY 512 MNGKRYSYKVELTHNTEFSYENSKNHLEKTLKALYMEDYSLRNIVKELKYKNL 571
Db 1010 I-----FNVKGIQENL--LGMFRSIE--TSIVIGSEKVDLN--1044
QY 572 SKIENEETLVENIKKDEQOLFEE--KITDEKPKDEKILEVSDYKVVY-----QKV 622
Db 1045 ---ENVVSSILDMINENKKEILNKLLENISSTEGVETVEHVBONYVVDVAPMDQFL 1101
QY 623 LNNKIDELKKQOLIKNVGLKHNHVPNSYKQEN-----KQPPYULIVLKEIDKL 674
Db 1102 GIINAGGLKEMFNLEDV-----FKSESQVITYEETIKDEPVQKEVEKEYSII 1150
QY 675 KVFMPRVESLINEEKNKIKTEGSDNSEPSTEGEITGQATTKPGQAGSALGEDSVQAQA 734
Db 1151 EEMEEINIVDLLEBEKEDL--TKDMIDAVEESIE-----ISSDS--KEET 1190
QY 735 QEOQKQAPVPVPPVPEAKAOPVTPPAPVNNKTENVSCLDYLEKLYEPLNTSYICHTIIV 794
Db 1191 ESIKDEKEDVSLVVEEVD-----NDMDESV-----EKVLELKN-----1224
QY 795 SHSTNMEKILKOYKITTKEESKLSGCDPLDLNLINQNNIPVMSMPSLNNLSOLFMEI 854
Db 1225 ---MEELMK-----DAVE-----INDITSKILEET 1247
QY 855 YE--KEWVCNLYKLDNDKIKNLLEAKKAVTSYKLTSSSMQPLSLTPQDKPVSANDPT 913
Db 1248 QELINEVADL--TKDMEKIKEL-----EKALSSED-----KEITIDAKDT 1285
QY 914 -----SH--STNLLNSLKFENILSLGKNKNIYOLIGOKSS--ENFEYKILKSDPTFY 963
Db 1286 LEKVIEEHHITTTLDEYVEL-----KYEVEKIKIEKVSDDLADLEDILKE-----1330
QY 964 NESFTNVSKADINSLNDESKRKLEEDINKLKTQLSFDLYNKYKIKLERLPDKK 1023

```



```

Db 1391 -----VKEIKELSEI--LEDYKELTIETDILEEKK--EIEKHDEKFEFEAEIKDLEA 1382
QY 1024 TWGKKMKQIKKLTLEKEDLESKLSNPNKHYLQNFVFNKKRAELAEFNTLENKI 1083
Db 1383 DI-----LKEVSSLEVEEKKLEVEHLEKE-----EVEHITSGD-- 1416
QY 1084 LKHVKGVLKYYNGESSPLKLTSESIOTEDNVALENFK--VLSKLGKLTNDNLLEKK 1141
Db 1417 --AHIKGL-----EEDDLEEVDDLKSGIIDLMLQDMELG--DMOKE 1453
QY 1142 KLVYSSGLHLHLAELEKVIKNNYTGNSPSENNTDVNNALSEYKKFLPEGTDAVTVSE 1201
Db 1454 SLEDVTKLGERVESLTKJVLISA--LGMDEQMKTRKKAQRKPLEVL-----LKEEYVE 1506
QY 1202 SGSDTLEBS-----QPKKPASTHGAGSNTTTTSONVDDEVDVYIPIPESEEDD- 1254
Db 1507 EPKKTKTKKVPKIDKEPEKDEIYEVEKMDIEDVEEDIEDIEDKVEDIDEDIDE 1566
QY 1255 DLGQVVTGEAVTPSVINDILSKIEENEVLYLKLPLAGYRSLSKOLENNVMTFNVNVDKI 1314
Db 1567 DIGE-----DKDEYIDILYQK--EKRIEKVAKK-----KLEKKVEBESVGLKHHVDEV 1614
QY 1315 LN--SREKFR--ENENKVNLESDDLIPYKDLTSSNYYVKKPYKFLNKKRDKFLSSYN-YIK 1369
Db 1615 MKYQKIDKEVDKESVKALESK-----NDVTN---VLKQNDQFSSVKV--NFKYKVKVFPA 1665
QY 1370 DSIDTDINADVIGY-----KILSEKYSKSDLSIKKYINDKOGENEKTYLPFLNN 1420
Db 1666 PFISAVAFASVYVGFESFSSCVTIASSYTL--LSKRVKTIN----KKKEPPFSYF 1718
QY 1421 IETIKTYNDKIDLEFVHLEAKVLYTEKSNVEYKIKELANVLTIOKIDAFKKNNNPV 1480
Db 1719 VFDFKNIK-----HTLOQMKKEFSK--EKNNNVI 1746
QY 1481 GIADLSTDYNNHNLTKFLSTGMWFEN 1507
Db 1747 EVTNKAEKKGNVQVTKTEKTKVDKN 1773

RESULT 12
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001

```

```

; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 4.2%; Score 350.5; DB 1; Length 2482;
Best Local Similarity 19.8%; Pred. No. 1.8e-10;
Matches 355; Conservative 281; Mismatches 617; Indels 537; Gaps 79;

QY 125 YADLKHR--VRNYLTITLKLKPOLFDLNNHLLTCDNIGHGKYLLIDG---YEELINELLY 179
Db 77 FSDQKHQKEIEMWCKLTSLT--GQVEDL--EHLQLLSN---ELMDQRCYQDLHAEYE 129
QY 180 KLNFFYFDLIRAKINDVCAND-----YCOITPFIKIRANELDVLKLVFGYRRPLDNI 231
Db 130 SLR---DLKSKDASLVNEDHQBSLAFDQOPAMHHSFAN-----IIGQGSMDSE 178
QY 232 KDNVGMEDYIKKNNKTIEINELIEESKKTIDKNNATKEEKKLYQAOYDSLTYNQ 291
Db 179 RSECRLEADQSPKNSAILQNRVDSLFSLES--OKOMNSDLOKCEBELVQIKGEIEENLMK 237
QY 292 LEEAH-NLISVLEKRIIDILKNN-----ENIKELDKINEIKNPPRANGTPTNLL 341
Db 238 AEOMHQSFVAETSQRIKIQEDTSAHQNVVAETLSALENKEKELO-----LL 284
QY 342 DKKKIEEHEKIKELIANTIKFNIDS-----LFTDPLELEYLRKNNKIDISAKETKE 396
Db 285 --NDKVFTEQAEIOELKSNHLLIEDSLKELQILSTLSL--KREMSITISLN--KREIFE 339
QY 397 STEPNEPNGVTPPLSYNDINNALNELNSFGDLINPFDTYKPSKNITTDNERKKFINEI 456
Db 340 LTOENG-----TLKEINASLNOEKW--NLIQ-----KSESPANYID--EREKSISEL 382
QY 457 KKKIKIER-----KRIESDKKSIEDRS---KSLNDTKEYEKILNLIYSKFPNNNDLINF 509
Db 383 SDQYKQEKLIILQRCETGNAVEDLSQYKKAQKNSKIECLNCTSLCEKRNKNEBOL 442
QY 510 EKMGMKRSYKVEKLTHTNTFASYENSK--HNLEKLTALK--YMEDYSLRNIVYKEKELKY 566
Db 443 KEAFKKEHQEFITKL-----AFAEERNQNMLELTVQALREKMDNDONNSSEAGGLKO 498
QY 567 YKNLSKIEINIEITLVENIKKDEOLFEKKTIKDE-----NKDEKTLLEVSDI-- 614
Db 499 EIMTLKEQNKMKQKQEVNDLQENBOLMKYKTRKHCQMLESPIRNSYKERESEBNOCNF 558
QY 615 ---VKVOYOKVLLMKIDELKKTOLILKNVLEK-----HNHV----- 649
Db 559 KPMQLEVAKEISLDSYMAQVQLEAMLRNKEKLEOESEREKCELOHELOTTIRGDLSTNL 618
QY 650 -----PNSYKQENKQEPYLLIVLKKKIDKLKVFMPKVESL 684
Db 619 QDMQSQELISGLKDEILDABEKYISGPHELSTQNNNAHLQCSLQTMKNKLN--LEKICEI 677
QY 685 INEEKKNITKESQSDNSPSTEGEITGOATTKPGQOAG-----SALGSDSYQ 731
Db 678 LQAEKYELVLTLELNDSEICIT-----ATRKMAEVEVKLLNEVXILLDDSGLLHGELVE 730
QY 732 ---AQAOEQQAQPPVVPVPEAKAQVPTPAPRNNKNTENYSKDIYLEKLYEPLNTSYI 787
Db 731 DIPGGEFQEPQNEQHPVSL-----APIDESNS-----YEHLLTS-- 764
QY 788 CHKYTLVSHSTNNEKIL---KQYITKEESKLS-----CDPLDLLEFNION 831
Db 765 -DKVEQMHFAELQEKFFLSQSEHKLLHDQHQOMSKMSKSELQTYVDSLKAENLVJSTNLRN 823
QY 832 -----NIPVWYSMFDSLANSLSQLEPMETIEKEMVCNLKYLKDNKYNL 875
Db 824 FQGLDVKEMQGLGELGVLPSLSSSCVDPSSSLSSIGDSSFYRAL---LEQTGDMKLSLNL 880

```



```

Db 1228 KL-OTENGELARQLEEKALIMQITRGKL-----SYTOQMEDLKROL---EEEGKA 1274
Oy 818 SS-----CDPLDLEFNIONNIPVYMSFMDLNNLSQLPMEIYEMVCNLK 865
Db 1275 KNAHLALASARHDCOLLBOYEEMEAKEALQRLVSKANSFVAQKRTYEDAIQRTIE 1334
Oy 866 LKD-NDKIKNLEAKKRVSTSVKTLSSSMQPLSTPQKPE--VSANDTSHSTNLN 921
Db 1335 LEEAKKKLAQRLODAEEAVEAVNA-KCSSLEKTKHRLQNEIEDLAWDVERSNAAALAK 1393
Oy 922 SLKLEFNILSLGKNKIYQELLQKSS-----ENFEKILKDDTQYNNES 966
Db 1394 KQRPNDKILAEWKQ--YEESQSELESQKEARSLTELFKLNAYEESLHLETKREN 1451
Oy 967 FTFNVKSKADINSLND-----ESKRKKLEEDINKLKTL 1001
Db 1452 ----KNLOEISDLTEQJGEGGKNVHELEKIRKOLEVEKLEQSLMEAEASLEHEBECK 1506
Oy 1002 ----QLSFDLYNKYKILER-LFDKKKTVGKYKM-QIKKTLLEKQLESKLSLNNPKHV 1055
Db 1507 ILRAQLEF--NOIKAEIRKLAKEDEMEQAKRNHLRVVDSLQTSILDATRNRNALRY 1563
Oy 1056 LQNFSEFFKKKEAEIAEENTLENTKILL---KHKKGVKYNNESSPLKLTSEESTQ 1111
Db 1564 ----KKMEGDLNEMEIQSOANRIASEAQKHLK-----NAQAH---LKDTQLO 1605
Oy 1112 TEDNVASLENFVYLSKLEKLDNLMLEKKKLSYSSGLHHLIAELKVKNNKNTYGNSP 1171
Db 1606 LDDAVANAND-----LKENIAIVERNRLLOALELELRAVYVEQERSKLAQEL 1655
Oy 1172 SENNDV-----NNALESYKFL-PEGTDVATVSESSGDTLEQSQPKKPASTHWGAES 1224
Db 1656 IETSEVQQLHSQNTSLINQKKKMDADLSQLOTEVEEAQOECRNAEEKAKKAITDAMMA 1715
Oy 1225 NITTSQ-----NVDEVDVITVIRFGESEBDYDQGV-VIGEATPVSIVIN 1272
Db 1716 EELKEQDTSALHERKKKMEQTIKDL-----QHRLEAQELALGGKKQLOKLEA 1766
Oy 1273 ILSKIENEVEVLYLKPLAGYVRSKQLENNVMTFVNVKDIINSFNRKKNVLESQ 1332
Db 1767 RYRELENELE-----AEQKRNAMESVGMKRSERIKELIYQTEDEKKN----- 1809
Oy 1333 LIPYKDLTSSNVYKDPYFLNKEKRDK--LSSYNYIKDSTID 1373
Db 1810 LVRLQDLVLDKLOLKVAKYKQAEAEQANTNLSKFRKYQHEID 1853

RESULT 15
US-09-104-324B-4
; Sequence 4, Application US/09104324B
; Patent No. 6232460
; GENERAL INFORMATION:
; APPLICANT: T rect, Ozlem, Sahin, Ugur; Pfreundschuh, Michael
; TITLE OF INVENTION: Methods for Diagnosis and Treating Cancers,
; TITLE OF INVENTION: And Methods for Identifying Pathogenic Markers in A Sample Of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski LLP
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,324B
; FILING DATE: 25-June-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: 08/892,702
; FILING DATE: 15-July-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6232460man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 318-3000
; TELEFAX: (212) 752-5958
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-09-104-324B-4

Query Match 3.9% Score 332; DB 4; Length 976;
Best Local Similarity 20.7%; Pred. No. 5.6e-10;
Matches 236; Conservative 204; Mismatches 354; Indels 344; Gaps 59;

Oy 196 CANDYQOIFP--NLKIRANEID--VLKKLVGYRRPLDINKDNYGKMEYIKKRNK-- 247
Db 41 CTFEDDLEFPRAKTNLSKNGENIDSDRALQKVN--LP--VLEQVNSDCHYQEGKLOS 94
Oy 248 TIEN-----INELLESKTIDKKNKATKEE-----EKKLYQAQ---YDLSI 287
Db 95 DIENSEGLRVFSFKLYKEEK-IKKKVSTFELRLROKESKLOENRKIIIEORKAIOELQF 153
Oy 288 YNK-----QLEAHNLLSVLEKRIQDITLKKNNIKELDKINEIKNPPRANGNPTNLLQK 343
Db 154 GNEKVSILKEEG-----IQENKDLIKENNAITRLCNLKE-----TCARSAEK 196
Oy 344 NKRIEHEKEIKELIAITFINIDSLFT-----DPLELYUREKNNKI-----D 387
Db 197 TKKYYEREETROYVMDNNINIEKMTTARQELAVQANSLNENHFKLDEYKELQNLQED 256
Oy 388 ISAKVETKES-----TERPEYNGVYTPSYNDINNALMELNSFGDILNPTQYTK 438
Db 257 YKKEINDKEQVSLDILQITEKENKKMDLFTL--ESRDKVQNL-----EKTU 305
Oy 439 PSKNITDNERKKFIMEIKIKIEKKKIESDKSYED--RSKSLNITREKELVNE 494
Db 306 QSENLKQSTEQHNLKLELDIVLSQRSYSTOKALEBDIQTKYTIQOLTKEKTOMER 365
Oy 495 IYDSKFNNDIDLTNFE-----KMGKRSYKVEKTLHNHTFASVENSKNH--LEKLTRA 546
Db 366 SNKARAHSPVYVEFEFTYQSLSEELLRTQQRLEKKNEDQKILTMELQKSSSELEMTKL 425
Oy 547 L-KYMEDYSLRNIVVEKELKYKNLISKIENIEFLVNIKQDEQLREKTKTDENKP 604
Db 426 TNKKEVELEELKLVLEKEFTLYEN-----KQEKIAEELKGTQEDQLGLQAREKEVN 479
Oy 605 DEKILEVSDIVKVOQVYLLMKNKIDELKQTL--LIKANVELKHNHVPNSYKQENKQRP 661
Db 480 D--LEI-QITAITTSEQYISKEVQD-KTELNEKILKNTLISHC--NKLSENKE-- 529
Oy 662 YLLIVLKEIKDLKVPMPYESLNEBKNIKITEGOSDNPSESTEGEITQATTKGQQA 721
Db 530 ----LTQETSDMTLELKNQOEDINNNKKQOEEMLKQIENLQETEQLRNE----- 575
Oy 722 GSALBGDSVQAQOAEQKQAPRVVPRPEAKQAVPPRPAPVNNKTENV-SKLDYLEKLY 780
Db 576 ----LEYVREELKQKR-----DEYKCKLDSSEENCNMLRKQOVENKKNYIEELQ 620
Oy 781 FLNYSYICHHYILVSHSTNNEKILKQYKIKTEESKLSQCDPLDLFNIONNIPVYSMF 840
Db 621 ----ENKALKKKGTAEKQLANVEI--KVNKLE--LEL-----LEL----- 650
Oy 841 DSLNNSLSQLFMEI--YEKEMVNCNLYLKNDKI--KNLLEAKKRVSTSVKTLSSSSMQ 895
Db 651 ----ESAKQKFGBITDQYKEI-----EDKKISEBNLEBEVK--AKYIAEAV-- 693
Oy 896 PLSLTPQDKREVSANDTSHSTNLNLSKLLENILSLGKNKIYQELLIGQKSSSENEFKI 955

```


Db 694 -----KLOKEIDKRCOH-----KIEMVALMEKHKQOYKIIERDSE----- 731
QY 956 LKDSOTFYNESFTNFKVSKADDINSLNDESKRKLKEEDINKKLTQLOSLFDLYNKYKTL 1015
Db 732 -----LGLYKKEOEOSL-----RASLEIESNLKAEI-----LSVKOLEI 769
QY 1016 ERLFKRKTVGKQKQIKLT-LKEOLESKSLNSLNPKHYLQNSVFENKKKEAIAET 1074
Db 770 ER--BEKE--KLKREAKENTATLEKKDK-----TOTF----- 799
QY 1075 ENLTENTKILKHXYGLVKKYNSPLKTLSEESIOTEDYNASLENFKVLSKLEGLKD 1134
Db 800 --LLETPEIYWK-----LDSKAVPSQTVSR-----NFTSYDH----- 829
QY 1135 NLNLEKKLSTL-SSGLHLHLAEKVIKKNKNTGNSPS-----ENNTDVNNALSEYK-- 1187
Db 830 --GISKDRDYIMTSAKKTLSTPLP-----KAYTKPTPKPLQORENINPIEESKKR 882
QY 1188 -----FLPECTOVATVYSESGSDTLEO--SQPKRPASTHVGAES-----NTITT 1229
Db 883 KMAPEFDINDSSSETTDLMSVSE--EETIKLYRNNNPASHLCVKTPKKAPSSLT 938

RESULT 16
US-09-134-001C-3242
; Sequence 3242, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NOCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3242
; LENGTH: 1151
; TYPE: PRM
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3242

Query March 3.78; Score 314.5; DB 4; Length 1151;
Best Local Similarity 19.78; Pred. No. 5.8e-09;
Matches 268; Conservative 225; Mismatches 485; Indels 379; Gaps 63;

QY 252 INELIE--ESKRTIDKKNNA---TKREEKKLYAOQYDLSYKQOLEAHNLISYLEK 304
Db 9 INEQDITLYKLKLEKESNNRPLINTINOYIKVILNOYTATFVGHSFGKSTLINLLE 68
QY 305 RIDTLKKNENIKELDKINEIKNPPANGTNTLLDKNKI-----EEHEKEIKELAK 359
Db 69 Q-DILP-----SSPVPTS-NTAIVSAKEDIILANLQOQYTKLKTND 111
QY 360 TIKFNIDSLFDPLELEYLREKKNKI-----DISAKVET-KESTERPEY-PNGVTPPL 411
Db 112 VKQNRQWVDEESIEINPSKFNGLFTFODTPGVDSNVATQOSTEOMFMYSNLFTYV 171
QY 412 SYDNINNALN-----ELNSFGDLINRPDYTKPEKSNITYDNRRKKEFINEIKIKIE--K 464
Db 172 DYNHQSALNFKMKRINEVGIP-----IFVINOIDKH-NE--EELTFETPK 216
QY 465 KRIESDKSYEDRSKSLNDITREYKELLINEIYDSKFNNNIDLTNEKMKGRYSYKVEKL 524
Db 217 SRVEKSIDW-----DI-----KLQDTYVYSKF----- 239
QY 525 THHNTFASYENSKHLEKLTALKTM-----EDYSLNIVYVEKELKYKMLIKIENE 577
Db 240 -----DHPQNEIDKLSNFLVMDQHRSTEDYVNRKTIOFIDVADQYI-----YONE 285

QY 578 IETIVENIKKDEOLPEKKITKDEKNPDEKILVEGSDIYVQYKYLMMKIDELKKTOLI 637
Db 286 MOSTIDTQIDNEQFEAVIYQFQNO-----EVSAREOLNDSQOLFNYLQOKR----- 335
QY 638 LKNVELKHNHVPNSYKQENKQEPYLLIVLKEIKRLKLVMPKVESLILNEEKNKITEGO 697
Db 336 ----DILDNAVIMTYDMRSLRN--YLESMAIDF-----KVGQFFKRRK----- 374
QY 698 SUNSEPTGEITGQATTKPGQAGSALGDS-----VQAOAQEQKQAPVPVPVPEBA 751
Db 375 KKEEDIKRLNEATYTOLOEAVNQVROPUREDMSPFLTRFKHAHVNEK----- 421
QY 752 KQOVTPPAPVNNKTEVNSKLDYLEKIEFLMTSYICHYILIVSSTWNEKILKQYKITK 811
Db 422 -----ILQEDVVPSL--ISELTO--TOTISNTYVL-----TSDEVIKALNKKI 464
QY 812 EESKILSSCDPLDLFNIONNIPVYSMEDSLNNSLSQLFMEIYKEWVCNLYKLDNDK 871
Db 465 EMES-----TPLEEAVNHQV-----NELSSDENEDRYEYDRYTELNTLKDSL 509
QY 872 IKNLLEAKKYSTVYKTLSSSMQPLSTPODKPREVSANDTSHSTNLNLSKLEENIIS 931
Db 510 SHNYKHYIHLDDSLDKLGRTEHFELKQENSTAYHRKHEFOHREFEVTSNQDIKRALD 569
QY 932 LGKNKIYQELIGOKSENFEYKILK-DSDFYNESFTNFKVSKADDINSLNDESKRKL 990
Db 570 IYKDVPLF-----DRTKODITITILRDNOITKVGFGTFSGKSSLINALLGDN---YL 621
QY 991 EEDIN-KIKKTLQLSFDLYNKYKTLERLFLPKKTVGKQKQIKKLTLLKEOLESKLSL 1049
Db 622 VSPNPTATATLSEYSGKESQITLK-----SKEQILLEVN-- 656
QY 1050 NNPKHYLQNSVFEPKPKK---EAEIATENTLENTKI-----LKHXYGLVAYYNG--ES 1099
Db 657 ---HVLEFEYLSFTNLFNFKISDLEKLSKLEKQOLAFSAIEKHY-----EMVYSMEH 708
QY 1100 SPLKTLSEESIQ---TEDYASLENFYVLSKLEGLKLDNLNENKKRLSYLSGLH----- 1151
Db 709 SLIHVYSLEIKKMSAEDYATF--VKTY-----HLKPLMDLKKKIILDISGLHSNNQR 761
QY 1152 ----HLAEKVIKKNKNTGNSPSENNTDVNNALSEYKFLPFGTDVATVYSESGSDT 1206
Db 762 HTNETEQILTSSDLLILVYVYFNHSFTDND-----KAFIEHMKD-----NO 802
QY 1207 LEQSQPKRPASTHVGAESITITTSQNVDEVDVYIIVPFGSEEDY--DOLGOV----- 1259
Db 803 LNEQAFK-----MIINAVDLAEKDODI-----QAVEDYADALGOVNLHSD 844
QY 1260 VTGEAVTPSVIDNLSKIENEYEVLYKPLAG---VYRSILKQLENVWTFVNVKDI-- 1314
Db 845 IYSVSSROSLSKNNNGINLRESIOYFAKVESBRTILDEQMTYQLOOMNNSFFKMKIDFND 904
QY 1315 ----LNSRPKNREKFN--VLESOLI PYKDLTSSNVVAKOPYKFLNKKERKDFLSSYNYI 1368
Db 905 DNAKLSARQKLNHYKNOTLILNQELIDPTAQTFNEVEBOYV-HNLERIKLQDLLE---V 960
QY 1369 KDSIDTIDINADVLYGYKILSEKYSQD-----DSIKKIYNDKGENEKYLP 1416
Db 961 KSVFNSQMTQNNDFNEEKISPKYIYLDQHQRLPLEQSLITERIKKYFNSQL--BEQILP 1018
QY 1417 FLNNIETLYKQVNDK-----IDLEFVILH-----AKVLYYKESNVEV 1455
Db 1019 VAKKLNQIHVITINAKFNVNPSLVYDALLQIETLSMQLSPKOLTKRKIYN---PKSQDI 1075
QY 1456 KIKELNV-LKTIQDKLADFKKNNNEVGIADLSTDYNH 1491
Db 1076 QVHIANQTELLQDDLSLRRQLN-----DYIYH 1103

RESULT 17
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C


```

; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820
; LENGTH: 1211
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

```

```

Query Match      3.7%; Score 310; DB 4; Length 1211;
Best Local Similarity 19.3%; Pred. No. 1.1e-08;
Matches 258; Conservative 233; Mismatches 471; Indels 374; Gaps 64;

```

```

QY 98 GSAVSGSGSRRTPSDN-----SSDSDAK-----YADLRKRVNLTITKELK 143
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 48 GVAIVGPNNGSGSNITDAIKWVLGQSAKSLGAKMEDITFGAEHRKAQNFAYEK-LK 106
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 144 YPOLFDLTNNHMLCDNINHGFKYLIDGYE-EINELLYK--LNFYFDLRAKINDV---- 195
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 -----LDN-HSQKQIDSEELVYVRRLYRSESEFYLLNDGARLKIDIDLF 151
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 196 ----CANDYQI-----PNNLK-IRANELDVLLKLVGYRK-----PLDN 230
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 152 LDGSLGKFAFSIISQGRVDEILNAKPIRRIIIEESAGVLYK-----YKKKASISQKLD 206
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 231 IKNVGGMEDYI-----KNNKK-----TIENINELIESKK 261
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 207 TENLNKVEDILYDLGSRVPLKEEAAIAKEYQOLSKEMQSDVYIVTSIDHYTEDNOR 266
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 TIDKNNATKEERKKLYQAQYDLSYKQLEEAHNLISVLEKRIDTLKKNENIKELLDK 321
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 267 LDERLNLHLKSOAKKEGQAOINQLOKYGKRRQONDYIEKLNLVELYKATENYEQLSGK 326
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 322 INEIKRPPRANGTPTLTDKNNKIEHEKEKEIKELAKTKFNIDSLFTPLLEIYLYRE 381
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 327 L-----NLYEERKKNOSETNARYEEELNDLESQIDSIKKNKAQNEKLLAD 371
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 382 -RKNKNDISAKVETKESTENEPNGVTYPLSTYNDINNALNELNSFGDLINPDYTKEPS 440
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 372 LKKKQOLNKEVOLES--LYISDQHDKELEIKNSYITLMSQSVYNNNDIRPLEHT 428
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 441 KNITVDNERKKF-----INEIKERIKIEKKIESDKSKYEDRSKSLNDITKEYEKLNET 495
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 429 IN---ENEAKRSRLDSRLVAFNQLKDIOQNIOTQKEYOSSKSMKVBQNIQOLBOQL 485
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 496 YDSKFNNDLTLNFEKMGKGRYSYKVEK-----THNTFASYNS--KHNLEKLTKA 546
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 486 TDSKRL--LSSEYENKLYQAYRYN-EKLSKRLDSLATQEDDYTFPFGVKKHLKANDKE 540
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 547 LKYMEDYSLNIVYVEKELKYKKNLISKIENIEFT-----LVENIKKDEBOLF-- 594
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 541 LR-----GHHGVAE-----VINPSEMTOAIEFALGASIQHVIYDN-EKDGQAIOYLK 589
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 595 -----KKITKDKNPKDEKILEV-SDIVVVOYQVLLMMKIDELK 633
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 590 ORGLGRATFLPLNWIDPRHVAADIKDVARSGOGFINASDAINVSAA--YONIEHLGL 646
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 634 TQILLANVELKHNIPVNSKQENKOEPLYLYLAKKIDELKAFMP-----KVESLINEE 688
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 647 NTVIVEN--LKHANLARAIRYRTR-----IVPLEGGVNVPGSGMGGARKKRSIISQK 699
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 689 KKNIKTEGSDNSEPTEGIEITQATTKPGQAGSALLESQVQAQOEOK-QAQPVPVPV 747
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 700 DELSTMNNOLEDYOROTA-----EPEKREKQKTOAE----- 721
QY 748 VPEAKAOPPPAPVNNKNTENV--KLDYLE-----KLYEF-LNTSYICHXYI 792
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 732 --QLESEFASOQYNNLKKQVHHHELDRLKTQEHNLKNEHEFFERKNDGQYQSK-- 787
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 793 LVSHSTWNEKILKOYKTKTEESKSLSSCDPLDLFNIONNIPVWYSFDSLNNLSOLF 832
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 788 --SKETLEK--ONHLEIOQOLK-----QLESDIERVTOJSKGAKSTHOTQ 832
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 853 EIEKEKAVCNLYLKNDKDKIKNILEAKKYSTSV-KTLSSSMQPLSLTQDAPREVSAND 911
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 833 QLHOKQ-----SDLAUVKERIKSQOYERLDKOLDSERKTI----- 870
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 912 DTSHTNLNNSLKLFEINISLGRNKNIYOBLIGOKSENFYERILKDSPTFYN-ESFTNF 970
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 871 -----EVNKKIKLPNS-----DEMGKDAFEKLEKQIQOQENRQMLNQGLSE 913
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 971 VKSKADDINSND--ESKRKLEEDT-----NKLKTLQLSFD--LYNKYKL 1013
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 914 IKOQRKDLNKEIKENESQLOKCHODILSIENHYQDIKAKOSKLDVLIHNAIDHLDNTYQL 973
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1014 KLERL--FDKRTVGYKKQIKRITLLEKQLES-----KLSLNPNKHVYNFNSVFNKK 1066
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 974 TVERARMEIDSDETIDNLR--KKVLTKMSIDELGPNVNLAIQEPBELNERT-PLNEQ 1029
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1067 KEAEIATEENTLENT-----KILKHKYLVR--YNGSESSPLKTLSEES 1109
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1030 R-TDLREAKETLEQIHEMKEVEGRKFTFHVAVODHFTTFVFKOLFQGGQAEIRL----- 1083
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1110 IOTEDNYSLENFVKLSLEGLKLDNINLEKKLSYSSGLHL--IAELKEYIKKNYI 1167
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1084 --TEDDLS--AGVDIYQPRK-----KLQHLISLSGGERALSALLFALLIKYR-- 1130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1168 GNSPSENNTVNNALDES-----YKKFLPE-GTVAVFVSSGSDTLEBQPKKPASTHYG 1221
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1131 -SAPFYLDEVAALDANVIRYAQYINELSTEQFIYIHRRGTMFSDBRLGVTYMQES 1189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1222 AESNTTTSQNVDEV 1237
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1190 GVSKLVSVNLTIDEV 1205
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 18
US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615

```



```

: REFERENCE/DOCKET NUMBER: 4600-0111.30
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 324-0880
: TELEFAX: (415) 324-0960
: INFORMATION FOR SEQ ID NO: 51:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1312 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NF.
: INDIVIDUAL ISOLATE: 389 TO 4324
: US-08-687-080-51

```

```

Query Match 3.7%; Score 308; DB 2; Length 1312;
Best Local Similarity 19.4%; Pred. No. 1.5e-08;
Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

```

```

Qy 411 LSYVDINNALNELSFGGLI--NPFDYTKPSKNITDNRKPKFINKIKIKIEKKKI 467
Db 84 LQFRDVN-----GELAVORSWCTQSKTKTEFTLGVTITRTGKGVSLSKCA 134
Qy 468 ESDKSYEDRSK--LNDITREYE-----KILNEYDSKFNNDLITNEKMMG 514
Db 135 EIDEMESSISGLVSKAVLNNTVFCHEODSNMPLSEKALKOKFDFLS-----A 182
Qy 515 KRYSYVEKLTHNTFASYSKHNLEKLTALYMEDY--SLRNIVYEK--LKYV 567
Db 183 TRYIALETTLRQVHOTQOCQKVEYQME-----LKYLKQYKKEACEIRQIISKEAQLTSS 237
Qy 568 KNLISKIENLETLENTKKEBOLFEKKTITKDEKPREKILEVSDIYKVOYK-----V 622
Db 238 KEIVKSYENELDPLKRNKLEIENHL--SKIMKLNE-----IKALDSRKKOMKDNSELE 290
Qy 623 LLMKKIDELKKTQILKVELKHNHIVNSYKOEKOEPPYLIYKKEIDKLYKMPKVE 682
Db 291 EKMEKVOGTDEOL-----NDLYHN--HQTIVREKERK-----LVQCHRELEKLN--KES 336
Qy 683 SLINEEKNKITEG-----QSDNSE-----PSTGEITGQ 712
Db 337 RLNOEKSELLVEQGRLOLQDRHDEHTRARDSLIQSLATQLDELDFGRGPFSEQIKNF 396
Qy 713 ATTPGQAGSA-----LEGDSVQAQAEQKQADPPVPPVPEAKQVPTTPAPYNNKTE 767
Db 397 HKLIVREOGEAKTANOLMDFAKETLKQKQ-----IDEIRDKTKTGIGRIIEKSE 448
Qy 768 NVSKLDVLEKLYEFLNTSYICHKYLIVSHSTWNEKILKQYKITEKEESKLSSCDPLDLF 827
Db 449 ILSK-----KONELKVN-----KTELQOLESSDRILEDELKAEKELSKAEK----- 493
Qy 828 NIQNNIPVMSMFDLSNLSLSQLEMEIYEKEMVCNLYKLDNDK-----IKNLEEAKK 881
Db 494 -----NSNVETLKMEVTSIQ-----NEKADLDRTLRLKLDQEKEQ 527
Qy 882 VSTSVKTLSSSQPLSTPDQKPEVSANDTSHSTNINSLKLFENILSL-----GNK 936
Db 528 LNH--HTTTRTQMELTKRDKADKDEQIRKIKSRHSDELTSILGIFPNKQOEDMLHSKSK 585
Qy 937 NIYO-----ELIGOKSENFYEKILKSDPTFYNESPTNFVSKAADINSLNDSEK 986
Db 586 EINTORRLAKLNLKELASSSEONKHNINELKKE-----EQLSVEDKLFYDGGSDOFFSD 641
Qy 987 RKLLEEDINKLK--TLQISFDLYNKYKYL-----ERLFDKKTGVGKYKMOI 1032
Db 642 LDRLEKEIEKSSQRAMLAGATAVYSQFITQUTDENQSCPCVQCQRFQTEMLQEVISDL 701
Qy 1033 K-KITLKEQLESKLSNINNKH-----VLQNFSVFFPKKKKAEALAELENTLENTKI 1083
Db 702 QSKRLAPDKLTSESELKKEKRDEMLGLVPRKOSII--DLKEKEIPDLRNKLQNVNR 759

```

```

Qy 1084 LKHVYGLVKKYNGGSESPKLTLSSESIQTEDNVASLENFKVLSKLEGKLDNLEKKL 1143
Db 760 DIQRLKNDIE---DEETLGLTIMPE---ESAKVCLTPTVITMERQOMELKD-----VEKI 809
Qy 1144 SYLSSGLHHLJAELEKVEIKRNWYTGNSPSENNTDVNNALESYKFLPEGTDAVAVYSESG 1203
Db 810 AQQAALQGL--LDRTRQVQYNOEKQKQHKLDIVSSKIELNRKLI----- 853
Qy 1204 SDTLQSQPKKPASTHVAESNTITTS-----QNVDEVDVYIYVIFGESEEDYDLDQV 1259
Db 854 QDOQOIOHLKSTNELSEKLOISTNLRQOLEEOT-----VELSTEVQSLVREIND- 907
Qy 1260 VTGEAVTPSIDNLSKIEENEVLYLKLPLAGVYSLKOLENNWTFNVWVKDILNSRF 1319
Db 908 -AKEVSP--LETLEKQOE-----KEELINKNTSNKTAQDLNLIK 948
Qy 1320 NKRENFKNVLESDDLIPYKDLTSSNYV--KDPYK-----PLNK-----EKDKFLS 1363
Db 949 EKVKNIHGYM-----KDI--ENVIQDKDYKKQKFTELNKVIAQLSECEKHEKINE 999
Qy 1364 SYNTIKDSIDP-----INFANDVL-----GYKILS-----EK 1392
Db 1000 DMRLMRQDIDPQKIOERNLQONLTLRKRNELEKVEEERKQHLKEMGQOYLQMKSEHOK 1059
Qy 1393 YKSDLSIKKYYINDKGENEKY-----LPFLNIIETLYK-----TVND 1430
Db 1060 LEENTDINKRNHNALAGQKQYEEBIIHFKEKELRPPQRRDAEYKREMIYARTTELYNK 1119
Qy 1431 KIDLFVHLEAKVNLVYTEKSNVEYKIKELNYLKTIOQLADFPKNNNF----- 1479
Db 1120 DLDIYYKTLQDAIMFH-----SKMKEIN--KIIRDLMSRYGQDIEYIEIRSDADE 1171
Qy 1480 -VGIAADLSTQYNNHNLTK 1497
Db 1172 NVASDKRRNRYRVVMLK 1190

```

RESULT 19

```

US-08-592-126-148
: Sequence 148, Application US/08592126
: Patent No. 5821091
: GENERAL INFORMATION:
: APPLICANT: Gregory Doljanov
: TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
: NUMBER OF SEQUENCES: 151
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Dehlinger & Associates
: STREET: 350 Cambridge Avenue, Suite 250
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/592,126
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Shultz, Charles K.
: REGISTRATION NUMBER: 38,615
: REFERENCE/DOCKET NUMBER: 4600-0111
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 324-0880
: TELEFAX: (415) 324-0960
: INFORMATION FOR SEQ ID NO: 148:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1312 amino acids

```


; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
 US-08-592-126-148

Query Match 3.6%; Score 307; DB 2; Length 1312;
 Best Local Similarity 19.4%; Pred. No. 1.7e-08;
 Matches 248; Conservative 214; Mismatches 453; Indels 364; Gaps 56;

QY 411 LSYNDINNALNELNSFGDIL---NPDYTKPSKNITYDNERRKFTNEIKIKIEKKI 467
 DB 84 LQFRDYN-----GELIAVQRSMVCTQKSKTEFKTEGVTYTKHGEKVSLSSKCA 134
 QY 468 EEDKSYEDRSKS---LNDITKEYE-----KLMEIYDSKNNNIDLTNFEKMG 514
 DB 135 EIDREMISSLGSKAVLNNVIFCHQEDSNMPLSEGRALKQKDEIRFS-----A 182
 QY 515 KRYSYVEKLTHTNTPASYSKSNLEKTLKALYMEDY-----SLRNIYKEK---LKY 567
 DB 183 TRYKALLETLRQVROGOKVKEYOME-----LKYKQYKERKCELRDQITSEADLTSS 237
 QY 568 KNLISKIENEIETLVENIKKDEBQLEFKKITDENKRPDEKILEVSDIYKVOOK-----V 622
 DB 238 KEIYKSEYENELDPLKNRLKEIEHNL--SKIMKLNE-----IKALDSRKQMEKDNSELE 290
 QY 623 LLMNKLDELKQQLILKNNVELKNHIVPNSYKQENQEPYLLVLKKEIDKLKVFMPKYE 682
 DB 291 EKMEKVFQGTDRQL---NDLYHN--HQRVYREKERK-----LVDCHELEKTLK---KES 336
 QY 683 SLINEKKNNIKTEG-----OSDNSE-----PSTEGEITGQ 712
 DB 337 RLINQKSELVLEQGRLOQADRHQEHIRAROSLIQSLATQLELDGFERGPFSEGRQIKNF 396
 QY 713 ATTKPGQOGSA-----LEGDSYQAQOEDCKAQAPRVPRVPAKQVPPRPAPVNNKTE 767
 DB 397 HKLYBRGEGEAKTANOLNNDFAKEETLKOQ-----IDIRKPKTGLGRHIEKSE 448
 QY 768 NYSKLDLYLEKLYEFLNTSYCHYLLVSHSTMEKILKQYKIKTEBESKLSGDDPLDLE 827
 DB 449 ILSK-----KONELKVN-----KYELOLEGSSDRILEDDQELIKARELSKAEK----- 493
 QY 828 NIONNIPVYMSFDLNSLSOLFMEIYEKEMVCLYKLDNDK-----IKNMLEAKK 881
 DB 494 -----NSNVELTKMEVISLQ-----NEKADLDRLRLKLDQEMEG 527
 QY 882 VSTSVKTLSSSSKQPLSLTPQDKPEVSANDTSHSTNLSNLSKLFENILSL-----GKNK 936
 DB 528 LNH--HTTTRTQWEMLTQKADQEDQIRIKSRHSDELTSILGYFPNKKQLEDMLSKSK 585
 QY 937 NTYQ-----ELIGQSSSENFEEKILKDSDTFYNESTFNVFSKADINSLDESK 966
 DB 586 EINQTDRLAKLNLKELASQONKNNHINNELKRE-----EQLSSYEKLDPVCSQDPESD 641
 QY 987 RKLLEDINKLKK---TLOLSPDLNKKYKLT-----ERLFDKKKTYGKYKMOI 1032
 DB 642 LPRKKEIEKSSKQRAMLAGATFAYVQSOFITQLTLDENQSCPCVQAFQVTEALQEVISDL 701
 QY 1033 K-KLTLLEQDLBSKLSLNNPKH-----VLQNSVFPPNKKKEALIAETENTLENTKI 1083
 DB 702 QSKTLRLAPKRLKSTESLELKKKERRDEMGLVPMROSII--DLKEKEIPELRKLNQVNVN 759
 QY 1084 LKHNYGLVKNYNGESSPLKLTSEESIQTEDNANSLNKKYVLSKGLKLDKNLNLKKKL 1143
 DB 760 DLORLNDE---EQETLLGTIMPEE---ESAKVCLDTVTIMRFQWELKD---VERKI 809
 QY 1144 SYLSGHLHLIAELKEVINKNKNTGNSPBNNDVNNALESYKRFPEGTDVATVASEG 1203
 DB 810 AQQAALQIGI--DLDRTVQOVNDEKQEHKLDTVSSKTELNKRL-----853
 QY 1204 SPTLEOSQPKKPASTHVGAESNTITTS---QNVDEVDVITVPIFGESEEDYDLDGOV 1259

DB 854 QDOQEQIQHLKSTYNELKSEKLEQISTNLOROOLEBOT-----VELSTEQSLYREIKD- 907
 QY 1260 VIGEAATPVVIDNIIISKINEYEVLYLKPLAGYRSLKQLENNVTFPVNNKDIANSRF 1319
 DB 908 -AKEQVSP--LETTLEKFOE-----KEELINKKTSKIAODKINDIK 948
 QY 1320 NKRENFKNVLESOLDIPKDLTSSNYV--KDPK-----FLNK-----EKRDPLS 1363
 DB 949 EYKKNHIGYM-----KQI--ENYIQDKDOVKQKQETELNNKLVIAQLSECEKREKIM 999
 QY 1364 SYNYIKDSIDTD-----INFANDVL-----GYTKIIS-----EK 1392
 DB 1000 DMRIMRODIDTQIKQERWLQDNLTLKRNBELKEVEERKQHLKENGQHQVLOMKSEHQ 1059
 QY 1393 YKSDLSIKKYINDQKGENKY-----LPLANNIETLYK-----TYND 1430
 DB 1060 LEEINIDIKRNHMLALGROKGYEEELIHFKKELREPOFQDAEKEKYEMAIWARTTELVAK 1119
 QY 1431 KIDLFYIHEAKVLANTYEKSNEYKIKELNYLKTQDKLADKKNNF----- 1479
 DB 1120 DDIYYKTLQALMKFH-----SMKMEELN--KIIRDLWRSTYRGODIEYIRISDADE 1171
 QY 1480 -VGADLSTDYNNHNLTK 1497
 DB 1172 NVSASDKRRNYRYVYMLK 1190

RESULT 20
 US-09-541-782-4
 ; Sequence 4; Application US/09541782
 ; Patent No. 6284480

; GENERAL INFORMATION:
 ; APPLICANT: Nislow, Corey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Beroud, Christophe
 ; TITLE OF INVENTION: Antifungal Assay
 ; FILE REFERENCE: 1015
 ; CURRENT APPLICATION NUMBER: US/09/541,782
 ; CURRENT FILING DATE: 2000-04-03
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 1038
 ; TYPE: PR1
 ; ORGANISM: *Saccharomyces cerevisiae*
 US-09-541-782-4

Query Match 3.6%; Score 303.5; DB 4; Length 1038;
 Best Local Similarity 21.2%; Pred. No. 2e-08;
 Matches 229; Conservative 169; Mismatches 395; Indels 287; Gaps 50;

QY 221 VFGYRRPLNINIKQNVG--KMEYIRKNNKTIENINELIESKKTIDKNNKATKEEEKKL 278
 DB 132 VEGRGSQDLIDPEVAGPLQDFIKGNCY-----LVYGMTST-----GKT 173
 QY 279 YQAQYDLSTYNNKQLEBAHMLI--SVLEKRIDTLK-----KNENIKELDKI 322
 DB 174 YTMGTGDEKLYNGELSDAAGIIPRVLLKLEDTLELQONDYVVKSFTELYNEBELKDLDD-- 231
 QY 323 NEIKNPPRANSQTPVTLDDKNNKIEHEKELEIKETIKFN-----DSLFTDP 372
 DB 232 -----SNSNSSTGTPD-----QOFMKLRIFFASSTANNNTSSASASSRSNSHNSP 278
 QY 373 LELE-----YLERKKNNIDISAKVETKESTEBNEVPNGVTYPLSYNDINNA---LN 421
 DB 279 RSLNDLTPRAALIRKLRLKSLPNTIKQIQOQOAVNSRNNSSN--SGSTTNASSNTN 336
 QY 422 ELNSFGDLINPDYTKPSKNITYDNERRKFTNEIKIKIEKKI-----KIESDK-KSYE 475
 DB 337 TNNQGSNAAPNDQT-----NGIYIQNLQEPHITNMAEGNLQGLKHHQVASTKANDPS 392
 QY 476 DRKSLNDITKEYEKLNEIYDSKFNNNIDLTNFERK-----M 513


```

Db 333 SRSHTIFPITTL-YKKHQBELFRISKMLVLDLAGSENINSGALNORAKAGSINOSLLTL 451
Qy 514 GKRSYKYEKLTHTNTFASYSKSNHLEKTLKALYMEDYSLRNIYVEKELKYNNLSK 573
Db 452 GRVINALVDKSGH-----IPFRES-----KLTRLQO-----DLSGN 483
Qy 574 IENEIETLVENIKKDEBOL-----FEKKTIDENKPKDEKILEVSDIVYQVKVLMKI 628
Db 484 TKTALIAITSPAKVTSEETCSTLEVASAKAKNNKPKOLGSFTMKDI-----LVKNIT 535
Qy 629 DELKKTQILKVELKHNHIV-PNSYKOEKNQEPYLLIVLKEIDKLVMFKVESLINE 687
Db 536 MELAKIKSDLLSTKSKREGIYMSQDHYKNLNSDLESY-----KNEVQECK---REISLSK 588
Qy 688 EKKNKTEGQSDNSEPSTEGEITGOATTKPGQAGSALGDSVOAQOAKOQAPVPVP 747
Db 589 NALLVKDKLKSKEITQSONCOIESLKTITL-----DHLRAQLDKQKHTLEI--- 634
Qy 748 VPEAKAQPPTPPAPVNNKTENVSKL-----DYLEKLYEPLTSTYCH--KYLVSHTM 799
Db 635 -----SDENNLOKLTETVOMALHDYKRRELD-LNOKFEMHITKEIKLKSTL 681
Qy 800 NEKILKQYKTKKEESKLSGCDPLDLFPNIONNIPVMSMPDLSLNSLSQLEFMEIYEREM 859
Db 682 FLQI-----NTMOQESILQET-----NIQPNL-----DMIKNEVLTLMRTMQEKAE 722
Qy 860 VCNLYKLDKNDKIKNLEAKKVTSTVKTLSLSSMQPLSLTPQDKPEVSAN--DDTSHT 917
Db 723 L--MYK-----DCVKILNESPKFENVV-----IEKIDILRVDPQKFKYKNAEMLSDIS 769
Qy 918 NLNNSLKLEFNILSGKN--KNIYOELIGOKSSENFYEKILKSDDTYNESEFTPVSK 974
Db 770 EENNMMKQYL-----KNHFFKNHQBELL-NRHVSTYENIEKRT-----NEVEVFKVYL 818
Qy 975 ADDINSLDESKRRKLEEDINKLKKTQILSPDLNKNYKLEKLEFLDKKKTGK-----YK 1029
Db 819 NDHL-----DENKKLIMHMLTTATSAVIDQEMDLFEPRKRVKENSFDLINDCDSMNEFYN 874
Qy 1030 MOIKKLTLLKBOLESKLSLNNPKHVLONFVFNNKKKEADIAETENT--LENTKILK 1086
Db 875 SMAATLSQIKSTVDTSSSMN-----ESISVM-----KGOVEESENISILKNNTKFND 923
Qy 1087 HYKGLVTKYNGESSPLK---TLSEESIOTED-----NYASLENFKVLSKEGLT 1132
Db 924 QFEDLINKNHMLKDNIKNSITSTSHITNVDDIYNTIENIMKNYGNKKNATKDEMIENIL 983
Qy 1133 KDNILNEKK---KLSYSSGJLHLLIAELKEVYKKNKNTGNSPSENNTDVNNALESYKFL 1189
Db 984 KEINLSKMPRLRLSNINSN-----SVQSYISPKKHAIEDENKSENVDN--EGSRKML 1035

```

RESULT 21

```

US-09-723-820-4
; Sequence 4, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1038
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-723-820-4

```

Query Match 3.6%; Score 303.5; DB 4; Length 1038;
 Best Local Similarity 21.2%; Pred. No. 2e-08;
 Matches 229; Conservative 169; Mismatches 399; Indels 287; Gaps 50;

```

Qy 221 VGYRRKPLDNIDKDNV--KMEYIKKNNKTJENINELLEESKRTIDKNNKATKEEEKKL 278
Db 132 VGPASQDLIPDEVAGPLFQFQIGYNGCTV-----LVYGMST-----GKT 173
Qy 279 YQAOYDLSTYKQLEEARHLI-SYLEKRIDLK-----KNENIKELDDKI 322
Db 174 YTMTEDEKLYNGLSDAGIIPRVLLKFLDLELOONDYVVKSPLELYNEELKDLDD-- 231
Qy 323 NEIKNPPANSNGPNTLLDKKKKLEEHKEKEIKETAKTIFNI-----DSLTFDP 372
Db 232 -----SNSNGSSNTGFD-----GQPMKILRTASSTANTTSNSASSSNSNRNNSP 278
Qy 373 LELE-----YLLREKNKNIDISAKVETKESTEPNEYPNGVTYPLSYNDINNA--LN 421
Db 279 RSLNDLTPKALLRKLRLKRLKSLPNTIKOOYQOQAVNSRNNSSN--SGSTNNMSSMTN 336
Qy 422 ELNFTGDLINPDYTKEPSKNIYTTNERKFFINETKEIKIEK-----KIESDK-KSYE 475
Db 337 TNNGQSSMAAPNDQY-----NGIYIONLOEFHITNAMEGLNLQKGLKHROYASTKMDPFS 392
Qy 476 DSKSLNDITKEYEKLELNDIYDSKFNNNIDLTNFEKM-----M 513
Db 333 SRSHTIFPITTL-YKKHQBELFRISKMLVLDLAGSENINSGALNORAKAGSINOSLLTL 451
Qy 514 GKRSYKYEKLTHTNTFASYSKSNHLEKTLKALYMEDYSLRNIYVEKELKYNNLSK 573
Db 452 GRVINALVDKSGH-----IPFRES-----KLTRLQO-----DLSGN 483
Qy 574 IENEIETLVENIKKDEBOL-----FEKKTIDENKPKDEKILEVSDIVYQVKVLMKI 628
Db 484 TKTALIAITSPAKVTSEETCSTLEVASAKAKNNKPKOLGSFTMKDI-----LVKNIT 535
Qy 629 DELKKTQILKVELKHNHIV-PNSYKOEKNQEPYLLIVLKEIDKLVMFKVESLINE 687
Db 536 MELAKIKSDLLSTKSKREGIYMSQDHYKNLNSDLESY-----KNEVQECK---REISLSK 588
Qy 688 EKKNKTEGQSDNSEPSTEGEITGOATTKPGQAGSALGDSVOAQOAKOQAPVPVP 747
Db 589 NALLVKDKLKSKEITQSONCOIESLKTITL-----DHLRAQLDKQKHTLEI--- 634
Qy 748 VPEAKAQPPTPPAPVNNKTENVSKL-----DYLEKLYEPLTSTYCH--KYLVSHTM 799
Db 635 -----SDENNLOKLTETVOMALHDYKRRELD-LNOKFEMHITKEIKLKSTL 681
Qy 800 NEKILKQYKTKKEESKLSGCDPLDLFPNIONNIPVMSMPDLSLNSLSQLEFMEIYEREM 859
Db 682 FLQI-----NTMOQESILQET-----NIQPNL-----DMIKNEVLTLMRTMQEKAE 722
Qy 860 VCNLYKLDKNDKIKNLEAKKVTSTVKTLSLSSMQPLSLTPQDKPEVSAN--DDTSHT 917
Db 819 NDHL-----DENKKLIMHMLTTATSAVIDQEMDLFEPRKRVKENSFDLINDCDSMNEFYN 874
Qy 918 NLNNSLKLEFNILSGKN--KNIYOELIGOKSSENFYEKILKSDDTYNESEFTPVSK 974
Db 770 EENNMMKQYL-----KNHFFKNHQBELL-NRHVSTYENIEKRT-----NEVEVFKVYL 818
Qy 975 ADDINSLDESKRRKLEEDINKLKKTQILSPDLNKNYKLEKLEFLDKKKTGK-----YK 1029
Db 723 L--MYK-----DCVKILNESPKFENVV-----IEKIDILRVDPQKFKYKNAEMLSDIS 769
Qy 1030 MOIKKLTLLKBOLESKLSLNNPKHVLONFVFNNKKKEADIAETENT--LENTKILK 1086
Db 875 SMAATLSQIKSTVDTSSSMN-----ESISVM-----KGOVEESENISILKNNTKFND 923
Qy 1087 HYKGLVTKYNGESSPLK---TLSEESIOTED-----NYASLENFKVLSKEGLT 1132
Db 924 QFEDLINKNHMLKDNIKNSITSTSHITNVDDIYNTIENIMKNYGNKKNATKDEMIENIL 983

```



```

Db 1186 SEE--QKRLNDIK-----LEISKSVSQYFMKLE--QKSVIGIKSVDFRNNELF 1236
Qy 1033 -----KKLVL-----LKEQESKLNSIN-----NPKHVLONF 1059
Db 1237 LSLPIONLTLPEMYRYFEMLYDIHIGIENKANREFIYSKSSINLDFLINDERVL-NL 1295
Qy 1060 SVFNNKKKEAIAETEMLEN-----IKILKHKGVLKYNNGSS 1100
Db 1296 EGLIKTKYLSLSLHRLTNSTSPADISIPLOITCPSITTIKK-----TEYYGHOLT 1350
Qy 1101 PLKLTSE-----ESTQEDNVASLENFKYLSKLEGLKDKNL----- 1136
Db 1351 NAMTVASVVRKPYDSNGLAINSIDK--SVSDVPALHTIVEBAKYNLLSMWDFYNTHASIM 1408
Qy 1137 -----NLE-----KKK-----LSYSSG-----LHLLAELKEVI 1161
Db 1409 DTIAROKHSTNIEFHPOSLLEDRDSKGLGLSLLYLDTGGYGGYQKLRHNIDTASTLY 1468
Qy 1162 KKNVYTGSPSENNTDVNNALSEYKKFLPEGTDVATVVSSEGSPTLEQSO-----PKKPA 1216
Db 1469 QTK-YNDNLKLSNDDF-----FLKTKQRIITMSNELGNRLKNAQOLEVELKDP 1518
Qy 1217 ST-----HVGAESNT-----TITS 1230
Db 1519 LLEGILVQRISLLITTEYHSLAQOISSFWRYTDPNFGHCDPFSLSLAQALFTIKNITSN 1578
Qy 1231 QNVDEVDVYIPIFGESEED--YDGLGOVYTGAV-----TPSYID-----NLSK 1276
Db 1579 RNFSLLYSGGLVKIYFESGLNNMRYIKLPVQTSLSLRLDIYLTPEKLISTSGSLINIGH 1638
Qy 1277 IENEYVLYLKPLAGVYR---SLKKOLENNVMTFNVN-----VK 1312
Db 1639 L---VPYSFIDIGVINGNINSESTVKKIKRSKINGDLQKHINHYLSEEGTQIK 1695
Qy 1313 DILNSRFNKRNFKNVLESDELIPYKDLTSSVYV---KDPYKFLNKKRDKPLSSYNYI 1368
Db 1696 DIVPFLGIQDNTIKYKLESIDIKPISEIQPLHSILSRQEHVKNLLSGLDDEFNKLKQ 1755
Qy 1369 KDSIDTQINFANDVLGYKKIKLSEKYSKSDLSIKKYINDKOENEKYLPFLNNIETLYTV 1428
Db 1756 GLSKTLYVLSVNF-----KESKINSIDVETVYTDLOGR-----LYXV- 1793
Qy 1429 NDKIDLVHLEAK-----VLNYTEKSNVEY---KIKELNYLKTQDCLADFKKNNFVG 1481
Db 1794 --DIDTVIGITFEKGINSLSEALEHNNIIDAINGVIGLYQYARMI-----KANDNISA 1844
Qy 1482 IADISTYNNHNLTKFLSTGWFENLAKTYLSNLLDGNLQGMNLISOHOCVK 1535
Db 1845 IDHAGAVSDIKNIYDKFLGILTLTNRVNPNPGVSGASLEGTSSGLEVCASR 1898

RESULT 23
US-08-290-919-3
Sequence 3, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DAREY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714637 CUSH
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label=X
OTHER INFORMATION: /note="X = M and N, or N"
US-08-290-919-3

Query Match 3.6%; Score 301; DB 1; Length 53;
Best Local Similarity 100.0%; Pred. No. 7e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1570 PNPICNENNGGCDADAKCTEDSGSNGKIKICECTKPPSYPLFOGIFSSSN 1621
Db 2 PNPICNENNGGCDADAKCTEDSGSNGKIKICECTKPPSYPLFOGIFSSSN 53

RESULT 24
US-08-480-604A-6
Sequence 6, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHAYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEY & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

```


APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Query Match 3.5%; Score 298.5; DB 1; Length 2710;
Best Local Similarity 20.4%; Pred. No. 1.2e-07;
Matches 339; Conservative 262; Mismatches 573; Indels 487; Gaps 90;

QY 84 SGGSVASGVSAGGVSAGGSGNSRRNPSSSDSAKSYADLKHRYANVLTITK--E 141
DB 538 TGGSLSEDDGVDENKNTALDKNYLNKKNIPSNVVEAGSKNYV-----HYIIQLQGD 590
QY 142 LKRPQFLD-----TNHMLTLCGNHGFK--YLDGYEELNELLYKLNFFYDILLRK 194
DB 591 ISTEATCNLFSPKPKNSITIIQRNMESAKSYFLSDGESTLEL----- 633
QY 195 VCANQYQIPEFNLIKIRANELDVLKLVFGY-----RKPLDNKDNVGMEDYIK 243
DB 634 ---NKY-RIPERLKNK-----EKVYVFIGHGKREFNTSEPARLSVDSLSNEISFLDTIK 685
QY 244 K--NKRTIE-----NINE-----LIESKKTIT-DKNKNATKEEKKKL 278
DB 686 LDISPKNVENVNLGCMNFSYDFVNEETYPGKLLSLMDKITSTPLPVPVKNKSI-----TI 739
QY 279 YQAOVYLSTYNNKOLEE--AHN--LISVLEKRIIDLTKNENIKELDKI--NEIKNP 333
DB 740 GANQYEVIRINSERKELAHSGWKINKEAIIJMSDLSKEYIT--FDSIDIKKL--AKS 793
QY 334 GNP-----NPLDKNNKIEHEKEIKIARTIKENIDSLFTDPLELEYILREKN 383
DB 794 KNIPGLASISEDIKTLIDLASV-----SPDTRKTIILNNKLINSSIGDYI--YY----- 840
QY 384 KNIDISAKYETKESTEPNEPNCVTPPLSLNDINNALNELNSGDLINPDIYKESKNI 443
DB 841 -----EKLEP-----VKNIH--NSIDDLIDELNLENSDEL 871
QY 444 YTDNEKRFINEIKIEKKIESDKKSYEDR-----SKSLNDI 484
DB 872 Y---ELKK-LNNIDEXYLISFEDISKNNSTYSVPFKNSGESVYVETEKEITSEHI 927
QY 485 TKYEKELNIYDSKFN--NIDLTFEKMGMGRYSYVEKELTHNTFASYSKHNIE 541
DB 928 TKRISTIKKSIITDVGNNLLDNTQDHTSGVNTLNMAFLQSL-----IDYSNNMDV 981
QY 542 KLTALK---YMEDYS--LKNIVYERELKLYKNNLISKIENE-----IETLV 582

DB 982 DLSTSVKVOLYQOLFSTGLNTIYDSIQL---VVLISNAVVDITNLVPTTEGIPVSTILL 1038
QY 583 ENIK-----KDEEQLEFKITKDEKPKDKILEVSDIVAVQYOKVL----- 623
DB 1039 DGINLGAIKELLEDHDPILKREL---EAKVGLAINMSTISIAATYASIVGIAEYTI 1095
QY 624 ----LMMKIDELKKTQILIKNVELNHNHVPNSYKOEKPEYVYLVLKKEIDKLKVE 678
DB 1096 LPYAGISAGIPSLVNNELIHD--KATSVVYFNHLSKSKYGP-----LKTEDD--KILV 1147
QY 679 PKVESLINEKKN-----ITEGOSDN-----SEPTSEELTGQATITP 717
DB 1148 PIDDLVISEIDFNNNSIKLQTCNIIAMEGSGFTVGNIDHFPSSPSISHTI-----P 1200
QY 718 GQOAGSALEDGSVOAOAOEOKAOPPVVPPVPAKAOV---PPPPAPVNNKTEN--VSKL 772
DB 1201 SLSTYSAIGIETENDPSKK-----IMMLPAPRSVFMEWEGAVPGLRSLENDGTRLL 1253
QY 773 DYLEKLY--EFLNTSYCHHYILVSHSTNNEKILKOYKITEEBSKLSSCDPLDLLENIQ 830
DB 1254 DSIRDLVPGKFWRFYAFEFYATITLKPVEDNTNIRIKLIDKOTRPFIM--PTITNEIR 1310
QY 831 NNIPWYMSFDSLANSLSQLFEMETYEKWCNLYKLKDNKIRNLLBEAKVSTSVKTL 880
DB 1311 NKL--SYS-FDGAGGTYS--LLSSY--PISTNNLSKDDLMIPNIDNEVREISIEGTTIK 1364
QY 891 SSSMOPSLSTLPQDKPEVSANDOTSHSTNLNLSLKLEPNITIS--LGKNNIYQELIGOKS 947
DB 1365 KG-----KLIKDVLSKIDINKKLI-----IGNOT 1389
QY 948 SENVEYKILKSDPTFYNESTNPFVKSKADDINSLANDESKRKLLEDINKLKTLOLSFDI 1007
DB 1390 ID--FGSDIDNKDRIYI-----FLTCELDKISLIIE-----INLAVKS----- 1425
QY 1008 YNKYKIKLELPKPKTKTVGKYKMOIKKLTLLKQLESKILNLPKHYLVONEVFPFNKK 1067
DB 1426 ---YSLLSG--DKNYLISMVNTIEKINTL--GLDS-----NIAVNYDESNNK 1470
QY 1068 EAEIETEMTLENTKILKLIY---KGLVYNGESSPLKTLSESIOTEDNYSLENFK 1123
DB 1471 FGAISKT-----SOKSIT--HYKDSKNILEFYV--DSTLEFNKSDPLAEDINVPMD-- 1520
QY 1124 VLSKLEGLKLDNLNLEK--KLSTYS-----SGH--HLIAELKEYIKKNYGNP 1171
DB 1521 --INTIGKYYVDNNTKSIDSFISLVSKOVKYNGLNESVYSYLDPEVKNDSGHNTS 1579
QY 1172 SENNTVNNALLESYKKEFLPEGTIVA-----TVYSEGSQDLE--QSQPKRPASHVGAESN 1225
DB 1580 NEMNLFLDN--ISFWKLEGFENINIEVDIKYFTLVGKTNLGVFEICONNKNIDILYFGEK 1638
QY 1226 TITTSQNVDEVDVILVLPFGSESEDDYDGLGVVGEAVTPVJINILSKINEVEVLY 1285
DB 1639 SSSKSTIFSGNGRVNVEEPIYND-----TGEDSTIS-----LDFSTE-- 1676
QY 1286 LKPLAGVYRSLLK--QLENNVMTFNVNVKD-----ILN--SRPNKRENFKNVLES 1331
DB 1677 --PLYGIDRYINKVILAPDLYTSLINININININININININININININININININ 1732
QY 1332 DLIPYK-----DLTSNNVYKDPYKFLNKEKDKFLS--SYNTIKDSID--TDINPANDV 1383
DB 1733 SSFEYKMWSTEGSDPLIVRVLSESNKKILQIRIKGLISTNQSNKMSIDPKDKIKLS--L 1790
QY 1384 GY-----YKILSEYKSDLSI-----KKYINDKOGENEKYLPLPN-----NI 1421
DB 1791 GYIMSNFKSNSNENELDRHGLGFIIDKNTYYVDEDSKLVKGLININNSLFTYDPIEFNL 1850
QY 1422 EPLKYTVNDKIDLVHLEKVLNYYEKSNSVEKIKELNLYKTIDOKLADFEKNNPVG 1481
DB 1851 VTGQOTINGKRYFYDINTGALTSYKI-----INCKHFFYNDGVQMOGVFGPGDGFEX 1904
QY 1482 IADLSTDYNNHNL-----LTKFLSTG--KVFENLAKTV 1512


```

Db 1201 SLSTYSAIGIETENIDFSKK-----IMLPLNAPSRYVMEETGAVPGLRLSLENDGTRL 1253
Qy 773 DYLEKLY--EPLNTSYICHKYLIVSHSTMNEKILKQYKTKREEKSLSCDPLDLFLNQ 830
Db 1254 DSIRDLPGKRYWYFAFDVAITTLKPYEDFTNIKIKLDKPTRNFM--PTITTNETR 1310
Qy 831 NNIPWYMSFPLSNLSQLEFMEIYEKEMVCLYKLDKNDKIKNLLEAKKYSTSVKTL 890
Db 1311 NKL--SYS-FGAGGTY-S-LLSSTY--PISTININSKODLMIFINDNEVREISENGTLK 1364
Qy 891 SSSWQPLSLTPDQKPEVSANDTSHSTMLNNSLKLEPNILS--LGNKNKYOEILIGOKS 947
Db 1365 KG-----KLIKDYLSKIDINKNKL----IGNOT 1389
Qy 948 SENFEKILKSDPTFNESEFTNFKSKADDINSLNDESKRKLLEDDIKKLTLOLSPDL 1007
Db 1390 ID--FSGIDKDKRY-----FLCELDKDLISLIE-----INLVAAS----- 1425
Qy 1008 YNKYKLELRLERDKKKRYGKYMOKIKLTLLEQLESKLNLSNPKHVLQNFSEFNKK 1067
Db 1426 ---YSLLSG--DKNYLSNLSNTEKINTL--GLDS-----KNLAYNTDESNNKY 1470
Qy 1068 EAEIETENTLENKILKHY--KGLVKYNGESSPLKLTSESIOTEDNYASLENFK 1123
Db 1471 FGALSKT-----SOKSII-HYKDKSKNILEFYN--DSTLEFNSKDFIAEDIVFMKDD-- 1520
Qy 1124 VLKLEGLKDLNLEK--KLSYLS-----SGLH--HIALEKYEIKNKNNTGNSP 1171
Db 1521 -INTTGKYYDNTDKSIDISILSVSKNOYKNGILNLESYSSYLDQVKNSDGHHNTS 1579
Qy 1172 SENNTDNNALLESYKFLPECTDVA--TVVSESGSDTLE--OSQPKPASTHVGASN 1225
Db 1580 NEMNLFLDN--ISFMKLFGEFENINFIIDKYFPLVGKTNIGYVEFIDGNKNNDIYGEKMT 1638
Qy 1226 TTTTSQNVDEVDYIYPIGSESEYDVLGQVVTGCAVPSYDNLISLKENYEVLV 1285
Db 1639 SSSKSTIFSGNRYNVEPIINPD-----TGEDIIST-----LDGSEY-- 1676
Qy 1286 LKPLAGVYRSLK-----QLENNVMTFNVNPKD-----ILN--SREKNREPNVLES 1331
Db 1677 --PLXGIDRIYINKYLIADLTSLNINTNYNSNEYPEIILYLANNTTHKKANNIN--LDS 1732
Qy 1332 DLIPYK-----DLTSSNYVYKDPYFLNKEKRDKFLS--SYNYIKDSID--TDINPANDV 1383
Db 1733 SSFEKMGTEGSDFLIVRYLESNNKILQIKIRKIGILSNTQSPNMSIDFKRIKLS--L 1790
Qy 1384 GY-----KYLSEKYSKSDLSI-----KTYINDKOGENEKYLPLN-----NI 1421
Db 1791 GYMSNFEKSPENSEMELDRDHGFKIIDNKTYYYDEDSKLVGLININNSLFFYDFIEFNL 1850
Qy 1422 ETLKTVNDKIDLFYIHLAEAVLNTYKESNVEYKIKELNLTAKTIQDLDLADFKKNNNVG 1481
Db 1851 VTGMOTLGGKYYFDINIGALTSYKT-----INGKHFYFNNDGVMOLGVFKGPDGEY 1904
Qy 1482 IADLSTDYNNNL-----LTKFLSTG--MVFNELAKTV 1512
Db 1905 FAPANT--QNNNIEGQALVYOSKFLTLNGKRYFDPNNSKAV 1943

RESULT 27
US-08-957-310-6
; Sequence 6, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200

```

```

; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-957-310-6

Query Match 3.5%; Score 298.5; DB 4; Length 2710;
Best Local Similarity 20.4%; Pred. No. 1.2e-07;
Matches 339; Conservative 262; Mismatches 573; Indels 487; Gaps 90;

Qy 84 SGGSVASGGSVAGSGVSGSGSRRTPSPDSSSDAKSYADLKHRYNLTLYIK--E 141
Db 538 TGGSLSEBNGVDVFNKNTALDKRYLLNNKIPSNVVEAGSKNY-----HIIQLQDD 590
Qy 142 LKYPDFL-----TNHMLTLCDNTHGFK--YLIDYEEINELLYKLNFPYDRLRAKLN 194
Db 591 ISEATCNLFKSNPKNSIIIOQNNMESAKSYFLSDGSIETL----- 633
Qy 195 VCANDYCOIPENLKRANELDVAKKLVGY-----RKPLDNKDNVKKMEDYIK 243
Db 634 ---NRY-RIPERLKNK---EKVKYTFIGHGKDEPNTSEFARLSYDLSLSNELSFLDIK 685
Qy 244 K--NKKTE-----NINE-----LIESKKTII--DKKNMATKEEKKKL 278
Db 686 LDISPKNVEVNLGCMNFSYDPNVEEYTPGKLLLSIMDKITSTLPDVKNNSI-----TI 739
Qy 279 YQAGYDLISYKQLEEE--AHN--LISVLEKRIIDLTKKNEIKELDKI--NEIKNPPANS 333
Db 740 GANQVEVAINSEGRKELLASHGKWINKREALMSDSKSEYI--FEDSDLNKTK-----AKS 793
Qy 334 GNTP-----NTLIDKNNKLEEHKEIKETAKTIKFNIDSLFTDPLLELYLRKN 383
Db 794 KNPGLASISEDIKILLDASV-----SPDRKFIINNKLNISSIGDYI--YV----- 840
Qy 384 KNIDISAVFERKESEPNEPENGVTYPLSYDNNALNELNSFGDLINPPTKPSKNI 443
Db 841 -----EKLEP-----VKNIH--NSIDLDIDEPNLEWVDEL 871
Qy 444 YTDNERKKFINKIKIKIEKKKIESDKSYEDR-----SKSLNDI 484

```



```

Db 872 Y---ELKK-LNNIDEXYLISFEDISKNNSTYSVRFINKSNGESVYETEKEIFESKSEHI 927
QY 485 TREKEXELNIDYKSFNN---NIDLTFNEKMKRYSYKEXELJHNTFFSYNSKHNLE 541
Db 928 TKETSTIKNSITIDVNGNLDNIQDLSOVNTLNAFFIQLS-----IDYSNNKQVLN 981
QY 542 KLTFRALK---YMEDYS--LBNIVAEKELKYYKLLISKIENE-----LETLV 582
Db 982 DLSTSVKVOYLADLFSTGLFTIYDSIOL---VNLISNAVDJLNLPTIEGPIYVSTIL 1038
QY 583 ENIK-----KDEQLFEKKITKDEKPKDEKILEVSDIVKVOQVL-----623
Db 1039 DGINLGAALKELDDEHPLKKEI---EAKVGLAIINMSLSIAATVASYIGAEVITFL 1095
QY 624 -----LNNKIDELKKTQULIKNVELKHNHVPVSYKOENQOEPRYLLVLKKEIDKLKLVF 678
Db 1096 LPLAGISAGIPLSVNNEELIHD-KATSVVYFPHLSKRYGP-----LKTEDD-KILV 1147
QY 679 PKVESLINEEKN-----IKTEGOSDN-----SEPSTEGELTGQATTKP 717
Db 1148 PIDDLVISEIDFNNNSIKLGTCLNLAEGSGHVTGNIDHFFSPSISHI-----P 1200
QY 718 GQAGASLBEDSVQAQOEKQAKOPPVVYVPEKAKY---PPRPVYNNKTEK--VSKL 772
Db 1201 SLSTYSAGIETENLDESKF-----IMLPNAPSKVFWETGAVPGLSLEDGTTRL 1253
QY 773 DYLEKLY--EFLNTSYICHYIIVSHSTNNEKILKQYKITEESKLSLSCDPLDLLENIQ 830
Db 1254 DSJRDLYPGKFFYRFAFFPYAITTLKPYVEDNPNIKIKLKDORNFIM---PITTTNEIR 1310
QY 831 NNIPVATSMDSLNNISJOLFMELYEKEMCNLYKLKDNKIKNLEAKKVSSTVTL 890
Db 1311 NKL--SYS--PDGAGGTYS--LLLSY--PISTNINLSKDDLMIFINDIVEIREISIENTIK 1364
QY 891 SSSMOPSLTRPOKPEVSANDTSHSTNLSNLSKLFENIIS--LGKNKNIYELQIGKS 947
Db 1365 KG-----KLIKDVLSKIDINKKLI-----JGNOT 1389
QY 948 SENFEYKILSDPTFYNESTFNEVKSADINSINLDESKRKKLEEDINKKTLQLSFDL 1007
Db 1390 ID--FSGDIDNKBRYI-----FLTCELDLSIIE-----INLVAKS-----1425
QY 1008 YNNYKLELLEFDDKKTAVGYKQKQIKKLLILKEDESKLSLNNPKVYLVONFSVFKKK 1067
Db 1426 ---YSLLSG--DKNYLISMLSMTIEKINFL--GLDS-----KNIAVNYDESNNKY 1470
QY 1068 EAELIETENLENTKILKLYH---KGLVYKYNESSPPLTSLSESIOTEDNVASLENFK 1123
Db 1471 FGAISKI-----SOKSII-HYKDKSKNILEFYN--DSTLEFNSKDFIAEDINVFMKD-- 1520
QY 1124 VLSKLEGLKLDNLEK---KLSYLS-----SGLH--HLIAELKEVIRKNNKYTNSP 1171
Db 1521 -IMTITKYYVNDNTDKSIFSLVSKNOVKVGLYLNESVSSYLDFAVKNSDGHHTS 1579
QY 1172 SENNTDVNNALLESYKFKLPBCTDVA-----TVVSESGSDLE-QSQKRRKASTVGAESN 1225
Db 1580 NFMNLFIDN--ISFWKLEGFENINEVIDKYFTLVGKTLGVEFTCDNNKNIDILYFGEWKT 1638
QY 1226 TITTSQWVDEVDVILVPIFGESEEDYDGLGVYKVTGEAVTPSYVIDNLSIKIENEYVLV 1285
Db 1639 SSSKSTIFSNGRNVVEPIYNP-----TGEDISTS-----LDSYE---1676
QY 1286 LKPLAGVYRSLK---OLENNVMPFNWVKD-----ILN-SFNNKRENEKNVLES 1331
Db 1677 --PLYGIDRIINKYLPDLTSLININTNYSNEYPEIIVLNPNFHHKVNIN--LDS 1732
QY 1332 DLIPYK-----DLTSSNVYVADPYKFLNKKRDKFLS--SYNTIKDSID--TDINFANDVL 1383
Db 1733 SSFEYKSTEGSPFILVRYLEESNKKILQKIRIGILSNQSFKMSIDFKDKIKLS--L 1790
QY 1384 GY-----KYIISEKYKSLDLSI-----KRYINDKOGENEKYLFLN-----NI 1421
Db 1791 GYIMSNKFSNSENELDRDHLGFKRIIDNKTYYYDEDESKLVGLTININNSLEFPDPIEFNL 1850

```

```

QY 1422 ETLKYTVNDKIDLEFVHLEKVLNTRYEKSNEVEKIKELNYLKTIODKLADFKNNFNVG 1481
Db 1851 VTMQOTINGKKKYFFDINTGALALSYKI-----INKKHFEFNDGVMQGLVGFPGPFEX 1904
QY 1482 IADLSTDYNNHL-----LTKFLSTG---WVEENLAKTV 1512
Db 1905 FAPANT--QNNNIEGQAIYQSKFLTLNGKKRYFEDNNSKAV 1943

RESULT 28
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oulne, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-323-170B-2

Query Match 3 5%; Score 298; DB 1; Length 3135;
Best Local Similarity 19.6%; Pred. No. 1.5e-07;
Matches 347; Conservative 270; Mismatches 575; Indels 582; Gaps 92;

QY 128 LKHVRVNYLLTIKE---LKYPOLFDLTNNMLTLCDNHGFKYLID---GYEINELLYK 180
Db 1502 LKNNLANF--TYKHQFYMYMILPALMDNDISFKCYVDLKKKKNVNSPLQPKYLRALYKK 1559
QY 181 LNFYFDLLRAKLDVNCAND---YCOIPFNLIKIRANELDVLLKLV--FGYRRPLDNIKDNV 235
Db 1560 LNIKFD-----NYVTGTDQNKYLTMYMDLHL-SHKKNYLKELFHDGKKKPAD-----1606
QY 236 GKMEDYIKKKKTIENINELIESKKTIDKNNKNA---TKREEKKKLYQAOYDL--SIYNN 290
Db 1607 -----TDANESII--ESLSINESNESGPPPTGDVDAEHLILEGYDTWESLYDE 1653
QY 291 QLEBA-HNLSVLEKRIIDTLAKKNENIKELDKINEIKNPPRPSNGTNPNTL-----LDKNN 345
Db 1654 QLEEVYINDIESLEK-----DIEQYVLYQVN-LKAPKLMSAQIHNRRHWCDPSKNN 1704

```


Qy	346	KIEHEKEIEIKIAKTIKFNIDS-LFIDPLELEYLREKKKNIDISAKV-----	392
Db	1705	LIVBESLKKKEBEGCNVPVNIHCYALLKRPIDLTLYKVCPTSKDVEAKVNISENDEYELQ	17644
Qy	393	-----EKESETEPNE-----YPNGV--TYPLSTNDINMLNELNSGDIINFD	434
Db	1765	VISLIEKRPHNFELIESKAKKNGDVVHNGVDGTGV-----LDNSFEEKYFKNIKIRK-D	1819
Qy	435	YTKPEKSNITYDNKRKKFVINE-----IKEKIKIEKK-----KI	467
Db	1820	KFEKPVIVNEDDTDEEKEDLESILPGAIIVSMVKLKKKKDDFTSYAAFPVPPYIPKDLHRKV	18797
Qy	468	ESDKSEYSSKSNL-----DIRKEVEKLLNEYDESKFNNNDINLPEKKM-GKRSY	518
Db	1880	ECANITEKDEKNOYISGNIHIDISNKRING--CDSSTNNSSLTSVSKVLNGEPEKN	1937
Qy	519	YKV-----EKLHNHTFASYENSKHNEKLTKLAKLYMEDSLNI	558
Db	1938	CEININNNVEFGIICDNETMLDEKCFHE-----IYSKDKVYKAKREVIAPNIDFESLNL-	1992
Qy	559	VEKEELKYK--NLISKI-----ENELTIVENIKKDEO-----	591
Db	1993	SNKKRVYAAVAPLDYIKLKLFFSCSCTSHNTIGTKAVTLNDEKEEBEDFTAGIKHNN	2052
Qy	592	-----LPEK-KITKDEKPRPEKILEVSDIYKVQVOKVLLMKNKIDELKTQOLL-----	638
Db	2053	VHLCNFPDNEBETDNNK-----IYCKITDAELFSEVILOIPRGT	20933
Qy	639	KNVELKHNHIVPNYSKOENKOPBYLLIVLKEKIDKLVEPMRVESLNEEKKNIKTGEO5	698
Db	2094	KNVE-----EGVONEEY-----KFKSLRPSLVFDNNNDIKVISC	21329
Qy	699	DNSEPTEGETGATTKPEQOQASALDEBQVQAQOEBQKQO-----PPVPVPVPAK	752
Db	2130	KN-----EVSISLAIK-----GYGNRIPTFKNGSKKEEGISFTIPLP-----	2167
Qy	753	AOVTPPAPVNNKTEENVSKLDYLEEPLNTSYIOCHKIYVSHSTMEKILKOYKITE	812
Db	2168	KQDTDLKFIINETIDNSNIKQGLIYIFVRKXVSENSFKLDFTGTSISMELNSQYKE	2266
Qy	813	EES--KLSDDPDLFLPNIONNIPWYMSMDS--LNSLSQLEPMELYEKEMVCNLYKIKD	868
Db	2227	KCKYVATIKKGD---JFGLK--CPKGFALFPQACSFNVLLVEYKSDYEDSEHINYIHKD	2280
Qy	869	---NDKIKTNLLEAKKVSYSVTKLTSSSQOPLSLTFPODKREVSADDDPSHNTNL-NSLK	924
Db	2281	KKYULKRRKDYIE-----LMDBNRRLO-NIQOYTGISNITDVLHAFENMLGMLP	2328
Qy	925	L-----FENISLGNK---KNIYOELIGOKSSBNFEYKILKDSOTFY	963
Db	2329	LNFKNHYSTAYAKVPDFNSIINFSCNCPYBEKHVYGTQVMSDBNRNF-DNIKKNENYIK	2387
Qy	964	NESFTNFVSKADINDINLINESKRRKI-----BEDINTKLTKLOLSFDLYNNKYLKL	1015
Db	2388	NELLPLNIEKAL---LLDBEROKKIKIOQOEBQOEBOLIKQODRLSRHDDYKKNHTYI	2443
Qy	1016	ERLFDKKRYGVKKYMOIKITLL--KEOLSKLNSLN-----NPKHV	1055
Db	2444	---LSDSMEHICDEYKENSLSITLPRNDTKIOKISICKINAKALADVUTIKCPHKNETPEY	2501
Qy	1056	LQNSVFNFKKKEAEIAETNLTENKILKHNKGLKYKYNNESSPLKTLSPESIQTEBN	1115
Db	2502	FPNSSLTINDOK-----IYTFDKKNEFYIIDPTK---KTFSLKDIYIOSF	2544
Qy	1116	Y-ASLENFVKYLSKLEGLKDNLNLKERRKLSYSSGLHNLIALKEVIKKNKNTGNSPEN	1174
Db	2545	YGVSLDHLNOIKKIHHEMWD-----VHLFPYPHNV--LHMVVLNNHI-----	2584
Qy	1175	NTDVNNMLLES--YKKFLPESTDVAIYVSSGSOTLEQOSPKKPASTHYGAESNITTTSQ	1232
Db	2565	-VNLSLSLEBGLVLEFKSKVYTDDETAT-----KKVTTLP--	2616

```

OY 1233 VDDEVDVITIIYPIGESEEDYDLAGOVYGEANT--PSYDILMSKIE-----NE 1280
Db 2617 --DGSSSLPIIPYVKE-----DIFHIFGCKSTTKKPNKAKTSLALIHIISSNFIING 2669
OY 1281 YEVEYLKPLGVGRYSKLEKOLENNVMTFENNVAKDIL-----NSRFRKRE 1333
Db 2670 CDFLYENQMDAISNNNNNSYSIFTHNKNTEMLICDISLIPKTVIGJGCKPNKKLANQOT 2729
OY 1324 NFKWV--LESLLPIYKULZSSNT-----VWADPKFLMKRKRDKFLSSYVI-- 1368
Db 2730 CFDEYVYKQEDVPSKTIYADKTNFESKDKIGNILKNAISINNPBKD--NTYTYLLLP 2786
OY 1369 ----KSDIPDINPANDVLGYKYLESEKYSKSDISIKKYINDKOGENE--KY-----LP 1416
Db 2787 EKFEELLIDTRKKVACACQDKKIYIIHMKIKSFMDKIK--IDCKKITIGDKICKYDVTYTKVA 2844
OY 1417 FLANNIEFLYKTV-----NDKT---DLEV--IHLKAVLVNY 1446
Db 2845 TCEJIDTIDSSVYLKEHHTTHVSYITLSRMDKLIIKYPTEKTEHFEFVWPNMLKDKVL-Y 2903
OY 1447 TYERS--WE-----VWIKELNYLAKTIOQDLAPFKNNPNVGIADLSPTYN 1490
Db 2904 NYNKPINIEHILPGAITDIYDTRIKIKO--YILRIPIPV---HKDIEF-----SLEFN 2932
OY 1491 HNNLITFEIStGVAFENIAKTVLSNLIDGNLOGLMNLISO----HQCQKOPONSGCFR 1545
Db 2953 NLSISLTK-QNONIITYGVAKI-----FIHINOYKEIHGC-----DETGRYS 2993
OY 1546 HLDRECKCCLLANYKQGBDKCVENPNPTCNENNG 1579
Db 2994 H-----LFTYSK--KPLPNDDICLVNITIG 3015

```

RESULT 29
 US-08-954-441-2
 Sequence 2, Application US/08954441
 Patent No. 6316000
 GENERAL INFORMATION:
 APPLICANT: Williamson, Kim C.
 APPLICANT: Kaslow, David C.
 TITLE OF INVENTION: Cloning and Expression of Plasmodium
 TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/954,441
 FILING DATE: 20-OCT-1997
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/323,170
 FILING DATE: 13-OCT-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/010,409
 FILING DATE: 29-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Einhorn, Gregory P.
 REGISTRATION NUMBER: 38,440
 REFERENCE/DOCKET NUMBER: 015280-113110US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 3135 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-954-441-2

Query Match 3.5%; Score 298; DB 4; Length 3135;
 Best Local Similarity 19.6%; Pred. No. 1.5e-07;
 Matches 347; Conservative 270; Mismatches 575; Indels 582; Gaps 92;

QY 128 LKHNRYNYLTITKE-----LKYPOLFDLTNMLTLCNINHFKKLID--GYEINELTK 180
 DB 1502 LKNNLNF--TYHQFNMYEIPALMDNDISFCIVDLKKKRYNVSPLGKYLALYKK 1559
 QY 181 LNFYDLAKLNDVAND--YCOIPFLKIRANELDYLKRLV--FGYKPLDNIKDNV 235
 DB 1560 LNKFPD-----NYVTGTOQNKLYMTYMDLHL-SHKRNLYKELFLHDLGKKKRPD----- 1606
 QY 236 GKMEYIKKNTKTIENINELIESKKTIDKNKA---TYEKKKLYQAOYDL--SIYNK 290
 DB 1607 -----TDANPESIT--ESLSINESNESGPPPTGVDAEHLLEGYDTWESLYDE 1653
 QY 291 QLEEA--HNLSVLEKRIIDLAKNENIKELLDKINELKNPPANGMPTNL-----LDKNK 345
 DB 1654 QLEEVLYNDLESLEK-----DIEQYVLOVN-LKAPLMSAOIHNNRHVCDPSKN 1704
 QY 346 KIEEHKEIKELAKTIKFIDS-LFTDPLELEYLYREKKNKIDISAKV----- 392
 DB 1705 LIYBESLAKKEELGKNPVNHCYALLKPLDTLYVKCPTSKDNEEAKVNSENDNEYELQ 1764
 QY 393 -----ETKESTEPNE---YPNGV--TYPLSTVDINNALNELNSFGDLNPPD 434
 DB 1765 VISLIEKRHFNEFTLESKRKRGNGVYVHNGVYDTGPV---LDNSTEKKYFKNIKIRP-D 1819
 QY 435 YRKPSKNITYTDNERKKFTINE-----IKERIKIEKK-----KI 467
 DB 1820 KPEFEKINIEDOTEEKDELSILPGALVSPMKYLLKKKDPPTSTAATVVPPIYVKDLHFKY 1879
 QY 468 ESDKSKYEDRSKSLN-----DITKEYEKLINETYDSKFNNNIDLTNEKKM--GKRRS 518
 DB 1880 ECNNTFEKDNQYISGNGIHIHIDISNMRKING--CDFSTNNSILTSVKLVNGETKN 1937
 QY 519 YKV-----EKLTHNTPASYSKSNHLEKTLKALKYMDYSLRNI 558
 DB 1938 CEININNEVEFGICDNETNLDPEKCFHE---IYSKDNKTYKRFREVIPIINDIFSLHN- 1992
 QY 559 VVEKELKYK--NLISKI-----ENETETLVENIKKDEO----- 591
 DB 1993 SNKKVAYAKVPLDIYINKLFCSCSCTSHNTITGTMKVTLNKDEKEDEDEKTAQGIKHN 2052
 QY 592 -----LFEK-KITDENKPDDEKILEVSDIYKVOVVLNKLIDELKKTOLIL----- 638
 DB 2053 VHLGNFQDPELFEFNK-----IYLCIKDAELFESVILQILFQGT 2093
 QY 639 KAVELKHNHIVNSYKOEKQEPYLYLVKKEIDKLKVPMPKVESLINEKKAIKTEGOS 698
 DB 2094 KNE-----EGVQNEEY-----KKSLEKPSLVEDNNNDIKVIYIGE 2129
 QY 699 DNSEPTGEITGOATTKPGQOQGSALLEGDSYQAOQEOQAO-----PPVPVPPPEAK 752
 DB 2130 KN-----EVSISLALK-----GYGNRIETFPKNGKKGGISFFLPI- 2167
 QY 753 AQVPPAPAVNNKTEVNSKIDYLEKLEFNTSYICHKYLIVSHSTMEKILKQYITKE 812
 DB 2168 -KODTDLKILINFTIONSNIKQGLIYFYKRVKVSNSFRLCOPFTGSGISLMLNQVKE 2226
 QY 813 EES--KLSSCDPLDLFNIGNNIPVMSFDS--LNNLSQLEFMEIYEREMCNLYKLD 868
 DB 2227 KCTCYIKIKGD---IFGLK--CPKGFALFPQACFSNVLEYYKSDYESEHINYIHKD 2280
 QY 869 ---NDKIKMLLEBAKKVSVTKLSSSQOPLSLTPQDKREVSANDTGSTYLN-NSLK 924

DB 2281 KRYNLKPKDVE-----LMDENPRELO-NIQOYTGISNTDVLHFKNNENIGLNP 2328
 QY 925 L-----FENILSLGN-----KNIVOEILIGOKSENFEKILKSDTFEY 963
 DB 2329 LNFKNHYSTAYAKVPDTPFISINFCSCNCPNPEKHVYGTQWVESDNNNF--DNIKKNNVYK 2387
 QY 964 NESFTNPKSKADINSLNDESKRRKL-----BEDINKLAKTYLQSLFDLYNKYKL 1015
 DB 2388 NFFLPIKEEYAL-----LLDDEERQKKIKOOQEEEOEQILKDDRLSHRDDNKNHNTYI 2443
 QY 1016 ERLFDKKTKYKQKQIKKLTLL-----KEOLESKLSLN-----NPKHY 1055
 DB 2444 ---LYDSNEHICYEKNESLSTLPNDTKIQAQSTCKIMAKALDVYTIKCPHRTKNTFPKDY 2501
 QY 1056 LQNFSVFNKKKEAEIETENLTENTLILKHKYGLVKKYNGESSPLKTLSEESTQEDN 1115
 DB 2502 FPNSSLITNDK-----IYATFDKKNFVYIDPFR--KTFSLKDIYIQGF 2544
 QY 1116 Y-ASLENFVYLSLECKLNDNLNLEKKLSYSSGHHLLAELKEYIKKNTGNSPSEN 1174
 DB 2545 YGVSLDLNOKIKKIHEDMD-----VHLFYPHNV--LANVLLNHHI----- 2584
 QY 1175 NFDVNNALES--YKFLPFGTQVATVVSSESDTLBQSPKRPASTHGAESNTTITSON 1232
 DB 2585 -VULSALBEGVLFMKSKVGTDEYAT-----KKNITLPT----- 2616
 QY 1233 VDDEVDVVIIVPIGESEEDYDGLQGVVTGEAVT--PSVIDNLISKE-----NE 1280
 DB 2617 --DGVSSILIPRYKE---DITFHLFCQKSTYTKPRKKNKTSIALIHIHISNRNIIHG 2669
 QY 1281 YEVLVYKPLAGVRSUKOLENNVMTFNVNKOIL-----NSRENKRE 1323
 DB 2670 CDFLYLENQTDNAISNNNNNSYSIFTHNNTENNLLICDISLPKTYIGIKCPKKKLNPOT 2729
 QY 1324 NFKNV--LESDDLRYKDLTSSNY-----VAKDPKFLNKEKRFELSSYNYI--- 1368
 DB 2730 CPEVYVYVQEDVPSTIYADKNYFESKDKIGILNALAISINNPDKD---NMYTYLLIP 2786
 QY 1369 ---KDSITDINFANDVLGYKILSEKYKSDLSIKKYINDKQENE--KY-----LP 1416
 DB 2787 EKFEELIDITKVKYLACTCNKYIIMKIKESTWDKIK--IDEKKTIGKIDCKYCTVYTKVA 2844
 QY 1417 FLNITETLYKTV-----NDKI---DLFY--IHLEAKVLNY 1446
 DB 2845 TCEIIDISSVLEKHNVHYSTILSRMDKLIKYPTNEKTHEENFVNPFLKDKVL-Y 2903
 QY 1447 TYEKS--NVE-----VKIKELNLYKTIQDKLADFKNNNFVGIADLSTDYN 1490
 DB 2904 NYNKPINIEHILPGATTTDIDYORTKIKO--YILRIIPRY--HKDIHF-----SLEEN 2952
 QY 1491 HNNLTKPLSTGNVFNLAITYVLSNLLDGNLQGLMINSQ-----HOCVYKQCPQNSGCFR 1545
 DB 2953 NSLSLTK--QNONIITYGNVAKI-----FIHINGYKEIHGC-----DFTGKYS 2993
 QY 1546 HLDERECCCLNRYKQBGDKVCNPNPTCNENN 1579
 DB 2994 H-----LFTYSK--KPLPNDDCNVITIG 3015

RESULT 30
 PCT-US93-07261-11
 ; Sequence 11, Application PC/TUS9307261
 ; GENERAL INFORMATION:
 ; TITLE OF INVENTION: P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREO
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John H. C. Blasdale
 ; STREET: One Giralda Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07940-1000
 ; COMPUTER READABLE FORM:


```

; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-07261-11

```

Query Match 3.5%; Score 296; DB 5; Length 1588;

Best Local Similarity 20.3%; Pred. No. 8.3e-08; Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

```

QY 243 KKKKTIENINE--LIEESKRTIDKNNATKEEKKKLYOAOYDLSTYKNOLEAHNLIS 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 12 EKNEKARNLKKKKLKEQKKNDQAKKDLTKKESQD-----SSSKSLKEKXNGA 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 301 VLEK-RIDPLK-----NEIKELDKINIKNPPRANSCNTPTLTLDKKKLEIEHEKEK 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 LKEKKEKELKLELENOKEKEKNKIKDNDEALKNKNDKDKKIVPKKESEYKDLK 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 EIAKTIKFIIDSLFTDPLLEYLYLRKKNK-IDISAKVETKSTENEPNGVTYPLSYN 414
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 EHELEKEKTIKQHLKDYEE-----RKRRNNWILRSLRKRLKLEIDELKLN----- 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 415 DINNALNELNSFGDLINPDYTKPEPSKNITDNERKKFTNEIKKIKIEKKIESDKSY 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 171 QLESAINELKERRASRPMMVKQRCMDDEVEMIKKYDDEQAKNGTDEEIKDKGDY 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 475 ED-----RSKSLNDITKEYEKLNITYSKFNNNIDLNFEMAKRRISYKEKLT 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 EIVETKRYGMREMLGEL-DEYEE-----RYEK--KRYLAK----- 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 HNTFASYSKHNLEKTLKALK-----YMEDYSLRNIYVEKELKYKMLISKIENEIEET 580
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 265 -----EDGEGDLKDVKEKLEETGYGFRKEFTTRILYKRR-----RKKEQK----- 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 581 LVENIKKDEEQLEFKKITKDENKPDKEKILEVSD-----IVKVOVOKVLLMNKIDELKKTQ 635
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 306 ---KLKEDEK-----EKKLIAEEDDEKIKLKDSDDKVVVNV----- 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 636 LLIKNEVLKHNHIVPSYQENKQEPYVL-----IVLKEKIDKLVMPKVESILNE 688
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 342 -----KNGSSFPDKFAPPKRTMYRLSELFPVPRKD-NELAVCGSDMSKVNVK 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 689 KKNIKTEGSDNSEPTEGEITGQATTKPGQOAGSALGDSVO-----AOA 734
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 393 KL-----KSTFNPFKRRRNKLEKRMKQELHFKKNKYKYKLE 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 735 QEQKQAPVPVPAKQVPTPAPVNNKTENVN-----KIDYLEKLEPLNTSYICH 789
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 432 REKRENPDEPLNTPEIHV---IRPEDLMDKGENSAGHPFKYQPTKGLKEY-EESHVSK 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 790 KYLVSHSTHME-----KIUKOYKITEEBSKLSLSCDPLDLFIQNNIPIPMYMPDS 842
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 488 DYQL-EHEPPTKLPEYKGVHSREYQLDHEPPTKLPEY-----KGVHSREYQLDNE 538
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 843 LNNLSQLEMEIYEKEMVCNLYKLDKNDKIKNLEBAKKVSTVKTLLSSSSMQPLTFQ 902
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 539 VADDELPE-----YEKGVHSREYQL-DNEGSPFLKEQDQTELAKGDITKKPHE--SVDEY 590
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 903 DKPEVSANDTSHSTNLNLSLKFENILSLGKNKI-----YOEL-----IGQ 945
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 591 DQTELAKGDITNKP--HESVDEYDQ-SELAKGKDITNKPHESVDEYDQTELAKGEYVN 647
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 946 KSENFYEKILKSDSTFTFNESTFNPKADINSINDESKK-----KLEEDINLIKKT- 1000
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 648 KPHEINLE--YNETDLAKGEYTNKPHESVDEYDQ-SELAKGKDITNKPHESVDEYDQTE 704
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1001 LDLSFDLYKRYKTLERLFDKKKTYGKYMQKILTLIKLEQLESKL---NSLNNPKHYLQ 1057
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 705 LAKGKEVTKAKADNLEFYNETDLAKGK-EYTKAKARENLEFYNETDLAKGEYTNKNAH--E 761
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1058 NFSVEFNKKKEADIAETENTLENTKILKHYGLVRYNGESSPLKTSIESIQEDNYA 1117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 762 NLEFY-----NETDLA-----KGKEVTNKAHNLNLEFYNETDLAKGEYTNKA-----HE 805
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1118 SLNKKVLSLKGKLLKDLN--NLEKKKLSYSSGHHILIALEKVIKKNKNTGNSPSN 1174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 806 NLEFYNETDLAKGEYTNKARENLEFYNETDLAKG-----KEYT-----NKAREN 850
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1175 NTDVNNALSEYKKE-LPEGTDVATVYSESGSDTLBOSQPKRPASTHVGASNTITTSQNV 1233
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 851 -----LEFYNETDLAKGEYTNKARENLEFYNETDLAKGEYTNKAREN--LEFYNE 900
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1234 DDEVDVITVPIFGESEEDYDLDGOVYGEAVTPSYIDNLSIKENY--EVLKPLIA 1290
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 901 TDLAKGEYTNKARENLEFYNET-DLAKGKEVTNKAARENLEFYEDKYAKNNELONKSGSD 959
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1291 GYVRSIK--KOLEN-----NWTFFVNVNKKDILNSFNKRNKN-----YL 1329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 960 GLKENALNKKELRNKSGDLKEMELNKKELRNKSGDLKENALNKKELRNKSGSEGLK 1019
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1330 ESDLIPIYKDL-TSSNRYVVDYPKFLNKKRKRDFLS-----YNYIKDSIDTDI 1376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1020 ENALNKKELRNKSGSEGLKENALNKKELRNKSGSEGLKENALNKKELRNKSGSEGLK 1079
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1377 NFANDVL--GYKKILSEKYSGLDLSIKKYINDKOGENEKYLPFLNIETLYKYVND-KI 1432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1080 ELKNEKLOKNG-----SEGLKENAEOKNKELONKSGSEGLKENALNKKELRNKSGDLKE 1134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1433 DLFVHLEAKVLTNYTEKSNVEYKIKEL-----NYLKT--TODKLADFK- 1474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1135 NMLKNNKELRNKSGDLKENALNKKELRNKSGSEGLKENVYTNNDKNDIONKDLNSKD 1194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1475 -KNNNFVGJADLSTDYNNHNLTKPFTSGMVEENLAKTYLSN--LLDGNLQGLNLSQHQ 1531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1195 MKNKKELNKNKDKMKNKELNKLDSN-----EDMKNNKELNKLKIDRNKDLKSGINNEQGN 1250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1532 CVYKQCP--ONSGCFRLHDERECKCLNTKQSGDKCEVNP-NPTCENNNGCCDADAK 1586
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1251 TGLKNTPKPGQONTGLKNTPNRQOONTGLKNTPESEOQNTGLKNTPESEOQNTGLKNTPN 1310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1587 CTEEDSG-----SNGKIT 1600
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1311 ERQOONTGLKNTPESEOQNT 1329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 31

PCT-US93-07261-16

; Sequence 16, Application PC/TUS9307261

; GENERAL INFORMATION:

; TITLE OF INVENTION: P1EMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREO

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: John H. C. Blasdale

; STREET: One Giralda Farms

; CITY: Madison

; STATE: New Jersey

; COUNTRY: USA


```

; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1663 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
; PCT-US93-07261-16

```

```

Query Match 3.5%; Score 296; DB 5; Length 1663;
Best Local Similarity 20.3%; Pred. No. 8.8e-08;
Matches 308; Conservative 248; Mismatches 601; Indels 362; Gaps 69;

```

```

QY 243 KKKKKTIENTEN--LIEESKTTIDKNNKATKEEKKKLYAQVDLSYINKOLEAHNLIS 300
DB 12 EKKEKRNALKEKKLEKQKNDQAKADLYKKESQD-----SSSEKLEKVNGEA 62
QY 301 VLEK-RIDTLK-----NENIKELLDKINEIKNPPANGTPTLLDKNKIEHEKEIK 355
DB 63 LKREKKEETLKKLEKQKKEKKNKIKONDEALKKNGDKDKKIVPKPPSVEKDK 122
QY 356 ELAKTIKFNIDSLFTDPLELEYLREKKN-IDISAKVETKESTEPNEYNGVYPLSYN 414
DB 123 EMELKKEFKIQHIDYEE---RKEKRRMWILRSLRDRLREIQELEKNA----- 170
QY 415 DINNALNELNSFGDLINPDYTKEPSKNITYTDNERKKFEINEIKIEKKKIESDOKSY 474
DB 171 QLESALINELKERASRRPMVMQKMGKDEVDWMIKKYDDEQAKNGTKDEELKDGDT 230
QY 475 ED-----RSKSINDITKEYEKLINLEIYDSKFNNNIDLTNPEKMMGRYSYVEKLT 525
DB 231 EEIVETKFGYGMRENALGEL-DEYEE-----RYEK---KRYLK----- 264
QY 526 HHHTFASYENSKINLEKTKALK-----YMEDYSLRNIYVEKELKYKNLISIEIEIT 560
DB 265 -----EDGEDLDLVEKLEETEGYFREKEPPTIRILVARK-----RNKEQK----- 305
QY 581 IVENIKKDEQLEFEKKITRKENKPKDEKILEVSD-----YKVOVQVYLLMKKIDELKKTQ 635
DB 306 ---KLKEDK---EKKLIAEPPDEKKIKLKSDSKVVPV----- 341
QY 636 LILKNEVELKHNHIVPNSYKQENKQEPYLL-----YLVKEIDKLKVPMPKVESLINEE 668
DB 342 -----KKSFPDKFRAPDKKRTMFYLSLFLYVPPKD-NELAVCGDSMDSKYNGK 392
QY 689 KKNIKTEGQSDNSEPSTEGEITGQATTKPGQAGSALGSDVQ-----AQA 734
DB 393 KL-----KSTFNPKRRRRNKLKERKMOELHKKKKNYKQKLE 431
QY 735 QEQKQKQPPVPPVPPAKQAVTPPPAPVNNKTENV-----KLDYLEKLYEFLNTSTICH 769
DB 432 REKRENDGEPILNTPETIHV---IRPSDLMDKGENKKSAGHPFKYQPTKGLKEY-EESHVSK 487

```

```

QY 790 KYILVSHSTNNE-----KILQYKITKEEESKLSSCDPLDLFNIGNNIPVYMSFDS 842
DB 488 DYOL-EHEPPTKLPEYKKGIVSREYQOLDHEPPTKLPEYE-----KGVHSREYOLDNE 538
QY 843 LNNLSQLFMEIYEKEMVCLYKIKONDKINLLEAKKAKSTSVKTLSSSMOPLSLITPQ 902
DB 539 VRDELPE-----YKGVHSREYOL-DNEGSPSTKEYDQTELAGKGDITNKPHE--SVDEX 590
QY 903 DKPEVSANDDTSHSTNLNLSLKLFEINLSLGNKNI-----YOEL-----IGQ 945
DB 591 DQTELAGKGDITNKP--HESVDEYDQ--SELAGKGDITNKHESVDEYDQTELAGKGEVTN 647
QY 946 KSSSENFYEKTLKSDPTFNYSFTNPKYSKADINSUNDESKR---KLEEDINKLKT- 1000
DB 648 KPHENLEE--YNETDLAKGKEVTNKPHEVDEYDQ--SELAGKGDITNKPHEVDEYDQTE 704
QY 1001 LQSLFDLNNKYKLELRLFKKTKVGYKNQIKLILKQLLESKL---NSLNNPKHYLD 1057
DB 705 LAKGKEVTNKAARENLEENETDLAKG--EVTNKAARENLEENETDLAKGKEVTNKAH--E 761
QY 1058 NFSVFENKKKEAEIETENTLENTKILKHYKGLVYKNGESSPLKTLSEESIQTEDNYA 1117
DB 762 NLEEY-----NETDLA-----KGEVTNKAHENLEENETDLAKGKEVTNKA-----HE 805
QY 1118 SLENFKVLSLEKELKDNLT--NLEKKLSYSSGLHILIAELKEYIKNNKYTGNSPSEN 1174
DB 806 NLEEYNETDLAKGKEVTNKAARENLEENETDLAKG-----KEYV-----NKAREN 850
QY 1175 NTDVNNALLESYKAF-LPEGIDVATVYSESGDPLLEOSQPKPKPASTHVAGASNTITTSQNY 1233
DB 851 -----LEEYNETDLAKGKEVTNKAARENLEENETDLAKGKEVTNKAAREN--LEEYNE 900
QY 1234 DDEVDDVYIIVPIRGSEBEDDLCGOVYTGAVPVSVIDNTLSIKTENY---EVLVYLKPLA 1290
DB 901 TDLAKGKEVTNKAARENLEENET--DLAKGKEVTNKAARENLEEYEDKYMKNELQNGSD 959
QY 1291 GYVRSLK--KQLEN-----NVMTFNVYVKDILNSRENKRENKN-----VL 1329
DB 960 GLKENNELKRELNRNGSDGLKENAELKNELNRNGSDGLKENAELKNELNRNGSGBLK 1019
QY 1330 ESDLIPTKDL-TSSNVYVKDPTKFLNKEKRDKLS-----YVITDSIDTDI 1376
DB 1020 ENAELKNELNRNGSGBLKENAELKNELNRNGSGBLKENAELKNELQNGSGEKENA 1079
QY 1377 NFANDVL---GYKILSEKKSDDLISIKYINDKQENKEYLPLFNNIETLYTVND-KI 1432
DB 1080 ELKNELQNGK-----SEGAKENABOKNNKELQNGSGBLKENAELKNELNRNGSGDLKE 1134
QY 1433 DLFVILHLEAKVLYTYEKSNEYKIREL-----NYLKT--IQDKLADF- 1474
DB 1135 NAEIKNELNRNGSGDLKEVAELKNELNRNGSGELKENYVTNNDLKNNDIQKDKLSNKO 1194
QY 1475 -KNNNFYGIADLSTDYNNHNLTKFLSTGVNPPENLAKTVLSN--LDDGNQGLNLNSQHQ 1531
DB 1195 MKNKEILNKDISKDKNNKELLNKDLISN---EDMKNELLNLDINKDKLDSIGNNEQON 1250
QY 1532 CVKQOCP---QNSGCFRHLDEREECKCLNLYKQEGDKCYVENPN-PTCNENNGCCADAK 1586
DB 1251 TGLKNTPPSKGQNTGLKNTPNEROQNTGLKNTPSGQOQNTGLKNTPSGQOQNTGLKNTPN 1310
QY 1587 CTEEDSG-----SNGKRT 1600
DB 1311 EROQNTGLKNTPSGQOQNT 1329

```

```

RESULT 32
US-09-572-191-2
; Sequence 2, Application US/09572191
; Patent No. 6355466
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman

```



```

: APPLICANT: Wood, Kenneth
: TITLE OF INVENTION: No. 6355466el motor proteins and methods for
: TITLE OF INVENTION: their use
: FILE REFERENCE: 1017
: CURRENT APPLICATION NUMBER: US/09/572,191
: CURRENT FILING DATE: 2000-05-17
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1388
: TYPE: PRT
: ORGANISM: Human
: US-09-572-191-2

Query Match          3 5%: Score 294; DB 4; Length 1388;
Best Local Similarity 19.4%; Pred. No. 9e-08;
Matches 29; Conservative 240; Mismatches 510; Indels 467; Gaps 70;

QY 93 SVASGASVAGSGGNSRRTNPSDSSDAKSYADLKHRYNNYLLTIKEL-KYPOLFDT 151
DB 207 SAARAYVLSGGMRRRVAASMSNRSRS-----HAV--FTTIESMEKSEIYVIR 257
QY 152 NMLTLCDNHGFKYLLDGE-----INELL-----YKLNFPDLRAKLNDVCAN 198
DB 258 TSLINLVD-LAGSEROKDTAEGRMLKEAGNINRSLGQVITALVDGNGKORHYCYR 316
QY 199 DYQIPIPNLKRAMELDVYKLVGYRRPDLNIDKNGKMDYIKKKKTI----- 249
DB 317 D-----SKLTF-----LNDSDG-----GNKTAIIANVHGS 344
QY 250 -----ENIELIEESKKTIDKNKATKEBEKKKLYOQYDLIYNKOLEE----- 294
DB 345 RCPBETISTLNFARAKLIKAKAVNEDQGNVQLAEVRLKEQLAEALSQGPRESF 404
QY 295 -----AHNLISYLEKRIDILKKNENIKE-LDDKINIKNPPRANGTPTLLDKNKI 347
DB 405 LTRDKKNTNYMEYQOAMLEFFKSEOEKSLIEKVTLED-----LTLKKEKF 452
QY 348 EEHEKEIKELAKTIKFNDSLF-TDPLELEY---YLREKKNKI--DISAKVET-KESTER 400
DB 453 IQSKMNI-----VKFREDDIIRLEKLHKESRGGLPEBQRLSELNENIOTLEQIE- 505
QY 401 NEYNGVYTLPSYNDINNALNELNSFGDLNPFDTKEPSKIYTDNERKKEFIEIKKI 460
DB 506 -HHRVAKYAME-----NHSLEENRRLRLPEVRAQE--MDAQTALEKAFSLISGME 558
QY 461 KIEKKKIESDKKSTEDSKSLNDITKEYEKLINELYDSKFNNDITLTFEEMGKRYSYK 520
DB 559 KSDKNQGFSPKQKEPLFAN--TEKLKQILOIOTELNNSKQYEFEKELTRR----- 612
QY 521 VEKLTNHTFASYENSKHNLEKL---TRALKYMEDYSLRNIVE-----KELKYYK 568
DB 613 --QLELESELOKANINLENLEATKACKROEYSQLNKHAETLKITTPRAYOYKIL 670
QY 569 NLISKIENETLVENIKKDEOLFERRITDENKPD--EKILEV--SDIVKVOYOKVL 624
DB 671 RPVFKLSPEMGSF--GSLYTQSSILNDILNEPVPREKMEQAFELISELTYOQMSAL 729
QY 625 MNKIDELKTKIOLIKNVELKHNIVPNSYKQENKQEPYLLIVLKKEIDKLKVPKYESL 684
DB 730 QAKLDE-----EHNK-----LKLQOHDKLEHNSHSTOMEL 760
QY 685 INEKKNIKTGQSDNSEPTEGETIGQATTKRQOQAGSALSGSVQAQADQOQADQRPV 744
DB 761 FSSERID-----WTKQOEELLSQL--VYLEKQOE----- 788
QY 745 PVPVPEAKAQVTPAPVNNKTEVNSKLDYLEKLEFPLNTSYICHYLLVSHSTNEKIL 804
DB 789 -----TQTKN-----DPLKSF--VHDLRVVLSHA----- 810
QY 805 KQYKITEESKSLSCDPLDLFNIONNIPVYGMFDSLNSLSQLEMEIYEKEVNCILY 864
DB 811 -----DKELSS-----VKLEYSSF--KTNQKEFKKLSERHMHVQL- 844

```

```

QY 865 KLKNDKIKNLLEKAKVSTSVKTLSSSQMPLSTTPQDKREVSAN-----DDTSHTN 918
DB 845 -----QDNLRLNEKEKLESKACLOSDYDNLQETIMKFEIDOLSNLONKKEKEDTLKSD 898
QY 919 LNSLKLFEENILSGKNKIYQOELIGOKSENEFEKILKSDTFYNESEFTNFKVSKADDI 978
DB 899 LNNLMELLE-----AKENNNKLSLOFEDKRENSKEILKYLEAVROEKQKTACE----- 950
QY 979 NSLDESKRRKLEEDINKKKTLOLSDLYNKKYKLEKLEFLDK-KTYVGKTKMOIKKTL 1037
DB 951 --QOMAKVQKLEESLATEKVIS-----SLEKSDSDKVVADLMNOIOE--- 993
QY 1038 LKQDESCLNSLNPKHYLQNFVFN-----KKKEAETENTLENTKITL 1084
DB 994 LRSSVCKETETIDYLKQELKINCKYNSALVDRRESRYLIKQEVYDILDKETLR----- 1048
QY 1085 LKHKGVLKYYNGESSPLKTLSEE--SIQEDNYASLENPKVL--SKLEGKLDNML 1138
DB 1049 -----LRIISEDIERDMLCEDLAHATEQJLMLTEASKKHSGLQSAQE 1091
QY 1139 EKKLSTYSGLHMLIAELKVIKKNYTGNSPSENNTDVNNALESTYKFLPEGTDAVY 1198
DB 1092 ELTKKREALIOELQKLNQKKEVEQK-----NEYNFKMQO-LEH 1130
QY 1199 VSESGDTEQSOQPKRPA--STHYGAESNTITTSQNVDEVDVYIYIFGESEDEDYDL 1256
DB 1131 VMSKAAE--DPOSKTPPHFQTH-----LAKLTLOEOEYID- 1165
QY 1257 GQVYVGEAVTPSVIDNLSKINENEVLYLKLPLAGVYSLKQLENNVMTFNVVKILN 1316
DB 1166 -----GRASKTS-LEHVTKLNEDREYKNAEL-----RKMEQL----- 1198
QY 1317 SRFKNRENFKNVLSDLIPIYDLTSSNIVYDPIYKFLKERRDXFLSYN-----IKDS 1371
DB 1199 --REMEMLR--LES-----QOLIEKMMLQOQ--LDDIKROKENSQNNPDMQOLKNE 1245
QY 1372 IDTDNFANDVLGYKILSE--KYSKSDLSIKKYINDQGENEYKLPNNIETL-YKTV 1428
DB 1246 QDESI--KERLASKIYEEMLKAKADEEVOYALYNKEMELKRTDTEVERTQTLIESAF 1302
QY 1429 NDKIDLFYHLEKAVLNTYKRSNVEYKIKELNLYKTIQDKLADPKKNNFVGLADLSTD 1488
DB 1303 QEKEDL-----RSKLEEMYEER--ERTSQEMEMLRKQVECLAE--ENGKLVG----- 1345
QY 1489 YNHNMLTK 1497
DB 1346 --HONLHOK 1352

RESULT 33
US-09-723-262-2
: Sequence 2, Application US/09723262
: Patent No. 6379912
: GENERAL INFORMATION:
: APPLICANT: Berard, Christophe
: APPLICANT: Sakowicz, Roman
: APPLICANT: Wood, Kenneth
: TITLE OF INVENTION: No. 6379912el motor proteins and methods for
: TITLE OF INVENTION: their use
: FILE REFERENCE: 1017
: CURRENT APPLICATION NUMBER: US/09/723,262
: CURRENT FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: US 09/572,191
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1388
: TYPE: PRT
: ORGANISM: Human
: US-09-723-262-2

```


Query Match	3.5%;	Score 294;	DB 4;	Length 1388;
Best Local Similarity	19.4%;	Pred. No. 9e-08;		
Matches 292;	Conservative 240;	Mismatches 510;	Indels 467;	Gaps 70

QY	93	SVGGGSAAGSGGNSRTRTPNSDSSDQAKSVADLKHVRYNLTILKEL-KYPLDPLLT	151
Db	207	SAAEAYOVLSGGMRRVAVASTSMNRESSR-----HAV-PTTISEMKSNEIIVNR	257
QY	152	NHMLTLCDNHGFYUULIDGYEE-----INELL-----YKLNFEYDLIRAKLNDVCAN	198
Db	258	TSILNTLVD-LAGSRQKDFHAEGRMLKEAGNINRSLSCIGVITLAVDVGNGKQHNHCVR	316
QY	199	DYCOIPFNLKIRANELLVYKLIVPGYKRPDLNIDKDNVGMEDYIKKNNKTI-----	249
Db	317	D-----SKLTF-----LIRDSIG-----GNAATIAIYANVHPS	344
QY	250	-----EINNELLEESKTTIDNNKNAKTEEEKKKLYOAOYDLSIYNKOLE-----	294
Db	345	RCGFETSLTNFAORAKILINKKAVNEDNEDGANSQLOAEYKRLKEDLALASGQTPRESF	404
QY	295	-----AHNLISYLEKRIDTLKRNENIKE-LDQINEIKNPPRANGSNTPTLIDKNKI	347
Db	405	LTRDKKNTNMYEYQOAMLEFPFKSEQEKSLIEKVTQLED-----LTLKKEF	452
QY	348	EEHKEIKELAKTIKENIDSLF-TDPLELEY-----YLRKNNKI--DISAKVET-KESP	400
Db	453	IQSKMM-----VKFREDOTIRLEKIHKSRCGFLEEDORLSLSELNEIOTLEQLE-	505
QY	401	NEYPNGVYPLSYNDINNALNELNSFGDILNPEDYKREPSKNIYTDNERKKIINEIKEI	460
Db	506	-HHBRVAKYAME-----NHSIREENRRRLRLEPYRAQE--MAQOTIAKLEKAFSEISGE	558
QY	461	KIEKKIESDKKSIEDSKSLNDITKEYEKYLEIYDSKFNNNIDLTNEKMMKGRYSK	520
Db	559	KSDNNQOGFSRKAKCECLPAN-TEKLYKOLLQIOTELNKSQOEYEEFKLTRRR---	612
QY	521	VEKLTHTNTFASYSNKHNEKL-----TKALKUMEDYSLNRIYE-----KELKYK	568
Db	613	-QLEISELQSIQKALNLENLLENTKACKRQEVSOQIMKIAETLIKITTPTRKAYOLHS	670
QY	569	NLSKINELEFTLVENIKKKEBOLFEEKTKIDENKD--EKILEV--SOIVAYOYOKVL	624
Db	671	RPVFKLSPEMGFS-GSLYTONSSILDNIDLINEVPPEMNOAFEAISELRTYOEMSAL	729
QY	625	MNKIDELTKQULIKANVELKHNIHVPRSYKOEKNOBOERYLYIJKKEDIKLYFMPRVSL	684
Db	730	QAKLDE-----EEHKN-----LKIQQVUDVLEHHSITQOBL	760
QY	685	INEEKNIKITEGOSDNSEPESTEGITGOATTKRGQOAGSALBEDSVQAOAOBOKOAPV	744
Db	761	FSERID-----WTKQOELLSQL--NVLEKQLE-----	788
QY	745	PVPPEKKAQVPTTPPAVNNKTKENVSKLDYLEKLYEFLNTSYICIKYLLVSHSTANEKL	804
Db	789	-----TOTKN-----DFLKSE--VHDLRVLHSA-----	810
QY	805	KOYKITEEESKLSOCPDLLENIQNNIPVMYSMPDSLNSSLSOLFMEIKEMYCNLY	864
Db	811	-----DKELSS-----VKLEYSF--KTQOEKFNKLSERHHVQL-	844
QY	865	KLKDNQIKINLEBAKVYSVYKTLSSSSMQPLSLTPQKPEVSAN-----DQTSHTN	918
Db	845	-----OLDNIRLENEKLTESKACLODSYDNLQEIKMFEIDLSRMLQNFKKENETLKSD	898
QY	919	LNNSLKLEFENLISGRKKNKIYQELIQKSSSENYEYKILKDSOTFYNESTNRYKSKADI	978
Db	899	LNNMELLE-----AEKERNKNSLQPEDEKENSKEILKVLAVRQEKETAKE---	950
QY	979	NSLNDSEKRLKLEBIDIKMLKTLQSLFDLYNKKYKLERLFK--KTVQYKMOQIKLTL	1037
Db	991	---QQMAKVQKLESSLATATEKYVIS-----SLEKSRBDSKKVAVADLMQOQE---	993
QY	1038	LKEOLESKLSLNNPKHVLONFSVFVN-----KKKEAEIATENTLENTKIL	1084

Db	Sequence	Score	DB	Length	Matches	Conservative
Db	994 LRSVCEKTEHIDTLKQGLKQKINCKYNALVDREESRVLKIKQEPVDDILDKETLR	1048				
QY	1085 LKHKGLVKKYNGESSPLKTISE--SIQEDNYSILENFVL--SKLEGKLDKQNL	1138				
Db	1049 -----LRITSEDERMDLQEDLAHAETQMLMTEASKKSHGLQSAOE	1091				
QY	1139 EKKKILSYSLGHLIAELKIVIKKKKNTGKSPSENNTDVNNALSTYKKPLPEGTDAVT	1198				
Db	1092 ELTKKALIQELQHKLNCKKEVEBKK-----NEYFKMKQ-LER-----	1130				
QY	1199 VSESGSDTLQESQPKPA--STHWGAESNTITTSQNVDEVDVLIIVIFGESEDDYDL	1256				
Db	1131 VMSDAE--DPQSPKTPHPQTH-----LAKILETQOEIED-	1165				
QY	1257 GQVVTGAEVTSVIDNLISKILENEYVLXKPLAGVYSLKQLENNMTFNVANKDILN	1316				
Db	1166 -----GRASKTS-LEHLVTKKINREYNNAELL-----RMKEQL-----	1198				
QY	1317 SRFNKRFRKFNVLSEDDLIPYDILTSNNVVRDPYFKLKKERDKFLSSYV-----IKDS	1371				
Db	1199 ---KEMELR--LES-----QQLLEKNMLLQO---LDDIRQKNSQGNHPDNOQLKNE	1245				
QY	1372 IDTDINPANDVLYGKYLSE--KYSKSDISIKKTYINDKQGENEKYLPFLNNIEL-YKTV	1428				
Db	1246 QEESI--KERLASKYIEMELKMKADLEVOASALYNKEMCLRPRTDVEERTQTLSEKAF	1302				
QY	1429 NDKIDLEVIHLAEKVLNTYKESNVEYKIKELNLYKTTQDLADPKKNNNFVGIADLSTD	1488				
Db	1303 QEKRL-----RSKLEMYEER--ERTSQEMEMLRKQVECLAE--ENGKLVG-----	1345				
QY	1489 YNNHNLTK 1497					
Db	1346 --HQLHOK 1352					
RESULT 34						
US-09-723-219-2						
Sequence 2, Application US/09723219						
Patent No. 6391613						
GENERAL INFORMATION:						
APPLICANT: Beraud, Christophe						
APPLICANT: Sakowicz, Roman						
APPLICANT: Wood, Kenneth						
TITLE OF INVENTION: No. 6391613el motor proteins and methods for						
TITLE OF INVENTION: their use						
FILE REFERENCE: 1017						
CURRENT APPLICATION NUMBER: US/09/723,219						
CURRENT FILING DATE: 2000-11-27						
PRIOR APPLICATION NUMBER: US 09/572,191						
PRIOR FILING DATE: 2000-05-17						
NUMBER OF SEQ ID NOS: 6						
SOFTWARE: FastSeq for Windows Version 4.0						
SEQ ID NO 2						
LENGTH: 1388						
TYPE: PRT						
ORGANISM: Human						
US-09-723-219-2						
Query Match						
Best Local Similarity 19.4%; Pred. No. 9e-08;						
Matches 292; Conservative 240; Mismatches 510; Indels 467; Gaps						
QY	93 SVASGGSVAAGSGGSRRTNPDSNDSQSDAKSYADLKHKRVNRYLLTIKEL-KYQDFDLT	151				
Db	207 SAALAYOVLASGGRRRRASTSMNBSRS-----HAY--FTITISMEKSEINIVIR	257				
QY	152 NHHMLTCDNNGFFKYLIDGYEE-----INELL-----KYLEFFYFDLRAKLADNYCAN	198				
Db	258 TSLNLTVD-LAGSEKQKQTHAEGRMKLEAGNINRSLQGLVATLVDVGNGKQKHVCYR	316				
QY	199 DYCOIPENLKITRANDELVLKTLVFGYRKRPDLNIDKQNGKMEQYIKKNNKT-----	249				

[illegible]


```

Db 317 D-----SKLTF-----LLRDSLG-----GNAKTAIIANVHPGS 344
Oy 250 -----ENINELIEESKTTIDKNNKATKEBEKKKLYAOYDLSTYNKOLE----- 294
Db 345 KCFGETLSTLNFPAORAKILKNAVANNEDTQGNVSOLQAEVRLKQLLELASGOTPPPSF 404
Oy 295 -----AHNLISVLEKRIIDTLKKNENIKE-LLDKINEIKNPPRANGTPTMLLDKNKI 347
Db 405 LTRDKKKNVMEYQOEAALFFKKSQOEKSLIEKVTQLED-----LTLKKKEKF 452
Oy 348 EBEHEKEIKAKTIKENIDSLF-TDPLELEY--YLREKNNKI--DISAKVET-KESTEP 400
Db 453 IQSNKMI-----VKFREDOJIRLEKIKHESRGFLPEOORLLSELNEIOTLLEQIE- 505
Oy 401 NEYPCGVTPPLSYNDINNALNELNSFGDLINPDYTKEPKSNKIYDNRKKRKINIKKI 460
Db 506 -HHRPVAKYAME--NHLSEENRRLRLLEPVKRAOE--MDAOTIALEKAFESISGME 558
Oy 461 KIEKKIESDKSYEDRSKSLNDITKEYEKLINELIYDSKFNNNIDLTNEKMMGRKYSYK 520
Db 559 KSDKNQOGFSPKAOKEPCLFAN--TEKLKAQLOIOTELNNSKQYEEFKELTRK----- 612
Oy 521 VEKLTHTNFASYSKNNLEKL--TKALKYMEDYSLRNIVF-----KELATYK 568
Db 613 --OLELESELQSKANLENLELATKACRQOEVSOQLNKIHAETLKITTPRTKAYQLHS 670
Oy 569 NLISKIEIEIETVENIKKDEEOLFEKKITKENDKP--EKILEV--SDIYKVOYQXVL 624
Db 671 RPKVKLSPEMSF--GSLTTONSSIINDILNEPVPENNEQFELISELRTVOQMSAL 729
Oy 625 MNKIDELKKTOLILKNVELKHNHIVNSYKQENKOEPRYLLIVLKKEIDKLKVFMPKVS 684
Db 730 QAKLDE-----BEHKN-----LKLQOHVXKLEHSHQOMEL 760
Oy 685 INEKKNIKTGQSDNSEPSTEGELTGOATTKPGQAGSALBGSVQAOBOKQOAPRV 744
Db 761 FSSERID-----WTQOEELLISOL--NVLKQLOE----- 788
Oy 745 PVPVPEAKQVTPPAPVNNKTENVSKIDYLEKLEFINTSYICKYILVSHSTWNEKIL 804
Db 789 -----TOTKN-----DPLKSE--VHDLKAVLHSA----- 810
Oy 805 KOYKITKEESKLSCDPLDLLFNIONNIPWYSMFDSILNLSOLFMEIYEKEMVCILY 864
Db 811 -----DKELSS-----VKLEYSF--KINOKEFKKLSRHHHYOL- 844
Oy 865 KIKNDKIKNLLEPAKKVSTYKTLSSSQOPLSLTPODKPEVSAN-----DTSHTN 918
Db 845 -----OLDNRLNEKELLESKACILOSIDYDNLQEIIMKFEIDOLSRHLQNFKENETLKSD 898
Oy 919 LNSIKLJFENLISGKNKNIYOELLIGOKSSENFYEKILKSDDTFYNEFTNFVSKADI 978
Db 899 LNNIMLELE-----AEKENNKLSLOFEEDKENSKEILLKLEAVROEKOKETAKE----- 950
Oy 979 NSLNDSEKRRKLEEDINKLKTLOLSPDLNKKYKILKLELFDK--KKTGYKKMOJKLTL 1037
Db 951 ---QOMAKVOGLLESILATEKVIS-----SLEKSDSDKKVADLMOQOE-- 993
Oy 1038 LKEOLESKLSLNNPKHVLQNFVFN-----KKKEAIEAETENTLENFKIL 1084
Db 994 LRSSVCEKTEFTIDTLKOLKIDINCKYNSALVDREESRVLKKEQVIDIDLKETTLR----- 1048
Oy 1085 LKHKGLVKKYNGESSPKLTSEE--SIQEDNVAJSLENNKVL--SKLEGLKLDNLNL 1138
Db 1049 -----LRLISEDELDMLCEDLAHAHQNLMLLEAKKSHGLQSOAOE 1091
Oy 1139 EKKRLSYLSGLIHLIAELKEVINKNKNTGNSPSENNTDVNNALLESYKFLPEGTDAVTV 1198
Db 1092 ELTKKALIOGLQHLKLNQKKEEVQOK-----NEYNFKMQ--LEH----- 1130
Oy 1199 VSESGDITLQESQPKRA--STHYGAESNTITTSQNDVDEVDDVIVIPFESBEDYDL 1256
Db 1131 VMSAAE--DPOSPTPPEHFOTH-----LAKLTLETOEOEIED- 1165

```

```

Oy 1257 GOVTVGEAVTPSVIDNIIKSIENEYEVLYLKLPLAGVRSKLRKLENNVFNWVNVKIDILN 1316
Db 1166 -----GRASKTS-LEHVTATKLEDEKVAENAL-----RKQOL----- 1198
Oy 1317 SRFNKNREKNVLESDDLIPYKDLTSSNYVVDKPYFLKNEKRDKFLSSYV-----IKDS 1371
Db 1199 ---REMENLNR--LES-----QOLIEKNMLQOQ---LDDIKRQKENSQNHDPDQOLKNE 1245
Oy 1372 IDTDINFANDVLGYKTIIE--KYKSDIDSIKTYINDKQGEKEKYLPLNNIETL-YKTV 1428
Db 1246 QEESI---KERLAKSKIYEEMLMKRADELEVOSALYKNEEMCBLRMTDDEVERTOLESKAF 1302
Oy 1429 NDKIDFVILHLEAKVLNVTYKESNVYKIKELNYLTIODKLADPKKNNPFVIADISTD 1488
Db 1303 QEKQOL-----RSKLEMYEER--ERTSQEMMLRKQVECLAE--ENGLYV----- 1345
Oy 1489 YNHNMLTK 1497
Db 1346 --HQLHOK 1352

RESULT 35
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match 3.5%; Score 293.5; DB 4; Length 1010;
Best Local Similarity 20.0%; Pred. No. 6.5e-08;
Matches 240; Conservative 201; Mismatches 417; Indels 341; Gaps 56;

Oy 226 KPLNDIKDNVGMEDYIKNNKKTIE---NINELIEESKKT-----IDKN 266
Db 3 KPLHIVMENFGPF-----IKETIDFEQVEVDQDLFLISGKIGSGKTMIFDAIVAYLYGMA 56
Oy 267 KNATKEE-----EKKKLYQAOYDLSIYNKOLEEAHNLISVLEKRIIDTLKKNENIKEL 318
Db 57 STKTRKEGDLRSHFADGSPSVIYQFVNNQTEFKIHREAPFIKGINTKTQAKLINIYEL 116
Oy 319 LDKINEIKNPPRANGNTPNLL-----DKNKKIEBEHEKEIKIATYI 361
Db 117 VDNQFELRE--SKVNOGNQFIVQLLGVAEQRQLEFILPDQGEKFFLOLSKDKOSILFTL 175
Oy 362 KENIDSLFTDPLELEYULREKKNKNDISAKVETKESTEPNEVPNGVTPPLSYNDINNALN 421
Db 176 -FNSFRPD-----ELRHVLVENKQEKV--QIENR-----YTQIENLMN 211
Oy 422 ELNSFGDLINPDYTKEPKSNKIYDNERKKF--INEIKEKIEKKIESDKSYEDRSK 479
Db 212 DIDFENN--DELAUYKE--LESSQDCKMIEKFPQFNQYDCKI-----LKSFE-- 255
Oy 480 SLNDITKEYEKLINITYSKFNNNIDLTNEFKMMGRKYSYVVEKLTHTNFASYSKSN 539
Db 256 AKNKTTEKELDOL-----NHKKYVNVLESENTRKL-----KAEKI---KPDOLKKEBOY 300
Oy 540 LEKTLKALKYMEDYSLRNIVVEKELKYKNL--ISKIENIEIETVENIKKDEEOL----- 592

```



```
DB 301 IDKLKOEKLMIOESKVL-----ITFTRLOGLSKKKODELVSLSHESQKLNFTNYHNEIK 353
QY 593 -PEKKTKITKENKPE-----KILEVSDIYKVQYQVILNMK-----IDELKK-----TOL 636
DB 354 GKOKOLEHSTRNEITTOFNQYLEKNQVFPNOJDKITSSYQOKPVYIEELKRLYSEYNDL 413
QY 637 ILKNVELKNIHVPNSYKOEKOEPEYLLIV-LKKEIDKLVKFMKVESLINEKKNIKTE 695
DB 414 ITRKEEL-----TKEMNNKKDFALLIEHYIEIYKLL-----KIIDSEBROKDE 458
QY 696 GOSDNEPSTEGEITGOATTKPGQAGSALGDSVOAQOEOQAPPVPPVPEAKQY 755
DB 459 KLEFDKQ-----LDKSSYLSKLEKKEKQOLNEISSITNIDATL 496
QY 756 PPPAPVNNKTEVSKLDYELKYEFLNLSYICHKY--LVSHSTNNEKLLKQYKITKEE 813
DB 497 ----IDLNDKKQVNE-----IKSAMSTIGDTCPIGNEIHSIGEHIIDESTIAQKNNKIKLE 549
QY 814 ESKLSSCDPLDLFNIONNIPVYMSFDSLNNLSLSQLFMEIYERKENCULYK-----L 866
DB 550 SKKV-----KIRDEIKIETRIEELNHNENELNEFEKQEKDISELQKQNLNHLNOL 599
QY 867 KD-----NDKIKNLEBAKVSTSVTLSSSQPLSL-----TPQCKPEVSAANDTSHS- 916
DB 600 KDOQOSINKLVENFEKOEKIVNKHOFD-----LDLSRKNQKREKLEIOINDFERHSQ 653
QY 917 -----TNLNSLKLFEINIISLGNKNKIYQELIG 944
DB 654 FSSVNDPEYIYSHAKKOVETYEYENKTKDKLMELNKKK-----IEMNQKILTEMLT- 707
QY 945 QKSSF--NFEYKILKDSDFPYNESFTNPKSKADINDSLNDESKRRKLEEDINKLKTKLO 1002
DB 708 QJSKEINNNELKMEKMOQLGFESEYDQ-VKSAA-DLSAQKDE-----IEREINIYNNK-- 758
QY 1003 LSRDLNKKYKLEKLEFDKKKTYGKKYKMOIKKTLLEKOLESKLNSLNKKNHVLQNSVF 1062
DB 759 -----YQSEIETNRL--KELYGKKLLNLEELROSTIEKTNLKLDETN-----799
QY 1063 FNKKRAEALTEENTLENTKILKHYKGLVYKNGESSPLK---TLESSEIOTEDNYASTL 1119
DB 800 -----SQATISYKIDONNSNKNFKIKNIIOIILDELKVKQKEIFELSELLAGNDVKRLTL 853
QY 1120 ENKRVYLSKLEGLKLDNLNLEKKKLSLSSGLHLIAELKLYT-----KKN 1164
DB 854 ENVYLLIYLE-KIIFQAN--QRLSEMSGNRYOLIR--RETIISLGLGELIDVDFEHSNK 907
QY 1165 NYNGNSPSENNTDVNNALSEYKKEFLPEGTDVATVY--SESGDPLEQSQPKKPASTHGAEL 1223
DB 908 SRIISLSGC-----ETFOASLALAGLSEVVOQESGCTILD-----SMFIDEG 951
QY 1224 SNTITTSQNVDEVDVITVPIFGESEBEDYDDGQVVTGEAVTPSYDNLISKIENEYE 1282
DB 952 FGLT-DQETLETAIDPLINKSSGR-----WGIISHVSELKORRPLLEVTLSNQYE 1002
```

RESULT 36
US-08-446-855A-2
Sequence 2, Application US/08446855A
Patent No. 5849573

GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA

```
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-855A-2

Query Match 3.5%; Score 293.5; DB 2; Length 2391;
Best local similarity 19.1%; Pred. No. 1.9e-07;
Matches 323; Conservative 252; Mismatches 543; Indels 577; Gaps 83;

QY 171 YEEIN-----ELAKLNFYDLRLAKINDVCANDYCOIPPNLKITRANEDLVKKIYF 222
DB 232 YKEINLEDPGNIDTLKYVCHNFIKRVI--KLN-----ITVNYK-----NKEEF 272
QY 223 GYKRPDLNIDKNGKEDYKRRKKTITENINLEIESKKTIDKKNKATKEEKKLYQAO 282
DB 273 NY---TNEKTTNDSMEDHNEINGISNPNNC--PSISSFDS-----ESKNV---- 316
QY 283 YDLISYKOLEAHNLISYEKRIIDPLKKNENIKELDKINEIKNPPANSQMPPTLTD 342
DB 317 ----INHTLRDKMNLITSSSEYELKL--HNCNFSNSSDK-----NDSEF 355
QY 343 KKKIEBEHEKEIKKIKTF--NIDSLFTDLELEYLREKKNKIDISAKYETKESTE 399
DB 356 KLYGIEYDYLIDLEENASFHYNVD-----ETGYVNNKNTNLSNKKIEQNNNE 408
QY 400 PNEYPGVATPLSYNDINNALNEL-----NSFGDLNPFDTYEPKNIYTD--NEKK 450
DB 409 NNR-----NKNKNNNNEVDYIKKDEDNNVNSKVFYSQYNNNAQNNHEETENLNN 457
QY 451 KFINKEIKIK-----IEKKIESDKSYEDRSKSLNDITREYKLEINEIYDSKFRNN 503
DB 458 DYSTYIRKKKKNEEFLNIVKRRVDHKEK-----IIVYIDCGIKNS 498
QY 504 IDLTNEKMMGKRYKYVEKLTNHNTPASE-----NSKNNLEKTLKALKYMEDYSLRN 558
DB 499 IIKNLIRHGMDLPLTYIIVPYUYNENHIDDAVLLSGPDRKPCDPLINKLSDSLTKN 558
QY 559 VV-----EKKELKYKNLISKIENEIETLVENIKRDEO-----LFEKK 596
DB 559 IIFGICLGNOLGISLSCDITKMKKTYGR---GVNQPIQIYVDNICYITTSNHHQCLKKKS 615
QY 597 ITRDE-----NKPEKILEVSD---IVKVOYQVLLNKKIDELKKTQLILKNVELKH 645
DB 616 ILKRELAISYINANDKSTIGISHKNGRFYSQFHP---EGNNGPEDTSFLKKNFL-- 669
QY 646 NIHPNSYKOEKOEPEYLLIVLKKET-----671
DB 670 --DIFNKKQRYRELYGINIYIKKKVLLSGGLICIGQAGEFDYSGTOAIRKSLKEGIYV 727
QY 672 -----DKIKVEMPK-----VESLINEKKN--IKTEGOSDNEPSTEGE 708
DB 728 ILVNPNIATVOTSKGLADKY-YFLPVNCFEVEKIIKKRKPDTILCTFG-----774
```



```

QY 709 ITGQATTPG-----QAQSALEGDSVOAQ-----AOEQOAPPVVPVPE 750
  |||
Db 775 --GQTALNCALMDQKVKLKNQCICLTSLIESIRITENRFLPAKLEINERI-APRGS 831
QY 751 AKAOVPPAPVNNKTENVSKDY--LEKLYEF--LNTSYI-----CHKYILVSH 796
  |||
Db 832 AK-----NNOAIDIANKIGYPLVTRTFSLGINSFIRNNEELIECKNIFL--- 880
QY 797 STAME-----KILK-----QKITKEES-----KLSSCPDLLEFNIONNIPWYSMD 841
  |||
Db 881 QTDNEIFIDSLQGWKEIEYELRDNKNNCIALCMMENIDPL-----GIHTGDSIVAVSQ 936
QY 842 SLNLSLSQLFMEIYEKEMV-----CNL-YKLKDNKIKRLBEAKKVSVTLSLSS 893
  |||
Db 937 TLSVYEYKREILAKYITHLNIIGECNIOFGINPOTGEYCIIEYNARLSHSLASKAT 996
QY 894 MQLP-----SLTPQDK-----PEVSANDTSHSTNLN 921
  |||
Db 997 GYPLAYISAKIALGYDLISLKNSTIKTTACFEPSLDYITTKIPRMDLKNREFASNTMNS 1056
QY 922 SKLFEIILSGKKNTYOELIGOKS-----SENFY-----EKIL-----K 957
  |||
Db 1057 SAKSVGEVMSIGRP--FEESI-OKSLRCIDNYLGFENYCIDWDEKKIIEELKNPSPK 1112
QY 958 DSDTFYNESTNPFYKSKAD-----INSLNDSKRRKLEE-----DINKLAK 999
  |||
Db 1113 RIDAIFHAFHLMPMADIHELTHIDYFPLHKEFYNIYMLQNKLYLKEQSLFNOLKPYFK 1172
QY 1000 -----TLQSFEDLYN-----KYKLERLDLKKTGKYKMQI 1032
  |||
Db 1173 HGFSQKQIAHYLSFNTSDNNNNNNNISCRTENDVMKRYRKL-GLEPHIKVIDTLSEF 1231
QY 1033 KKL------LKEQLESKLSLNPKEHVLQNFVFPNKK-----EAEI 1071
  |||
Db 1232 PALTNLYLYTQOEHVDYLPNMMKRIKICLNNKRNA-----NKKYHVKNHLYNEV 1283
QY 1072 AETEN-----LENTKILKHKGVLKYVYESSPKTSESIOTEDN 1115
  |||
Db 1284 VDKQDLOLHKNNNNNNNNGVENEKCKLNKESG--YNNSSCINT--NNINIEEN 1336
QY 1116 --YASLENFYV-LSKLEKLDKNLMEKKKLSYLS--SGLHLIAELKEVI----- 1161
  |||
Db 1337 ICHDISINKNIKVTINSSNISNENVE--TNLCVSEBAGSHIYGEKESIGSDDTNI 1395
QY 1162 -----KKNRTGNSPSEN--NTDYNNALLESYKKFLPGCTDAIYVSESQDTLQSOQPK 1214
  |||
Db 1396 LSAONSNNNNSCENNENKKNADVNVLENDTK--REDINTTIVFME-GQNSVINNNKE 1452
QY 1215 PASTHVAESNTITTSQVNDDEVVDVILVPLFGESEEDYDOLGOVVGAVTPSVJINIL 1274
  |||
Db 1453 NSSLTKDEDEDIYVNVNKKENNTNSVI-----NNVDCRKKDMQK--NINDECK 1499
QY 1275 SKIENEYEVLYLRPLAGVYSLKQLENNVM-----TFNVNVKDLNLSR-----F 1319
  |||
Db 1500 TYRKNKRYKDM-----GLNNNIYDELNGSHSTNDHLIDNFNTSDEIGNKNKMDMYL 1553
QY 1320 NKNENFNVLESDLIPKQDLSNYYVVKDPRKFLNKEKROKFLSSYNTIKDSITDIN-- 1377
  |||
Db 1554 SKESKISN-----KNPGNSYYVDSVYNNMEYKINKKKELDENLDEYNANNVMN 1604
QY 1378 -----FANDVIAGYKILISEKYSKDSIKKY-----INDKOGENEKYLPFLNIEFLY 1425
  |||
Db 1605 CSNTNNAASAVVNGKDRNDLENDCIEKNMDHTYKHYNLRNNRSTNERMMLAVNN----- 1659
QY 1426 KYVNDKIDLEFVILHLEAVLNT--YEKSNEYEVKIKELNLYLKTIDOKLADFKKNNFVGA 1483
  |||
Db 1660 -----EKESNHEKGRHNGLNKKNKEKMEKN--KGNKMD--KKNYHAYN-H 1701
QY 1484 DLSTDYNNHNLTKF 1498
  |||
Db 1702 KRNNEYNSNIESKF 1716

```

```

RESULT 37
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 618396
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 618396
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: A093/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ. ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Query Match
Best Local Similarity 19.18; Pred. No. 1,9e-07;
Matches 323; Conservative 252; Mismatches 543; Indels 577; Gaps 83;

QY 171 YEEIN-----ELLYKLFYFDLRAKLNDVCANDYCOIPEULKIRANELDVLKLVF 222
  |||
Db 232 YKEINLPDPCNIDILKVCVNHFIYI--KLNN-----ITYNKK-----NKEEF 272
QY 223 GYRPLNINIKNVGMKMDYIKKNTJENINELIESSKTTIDKKNNTKBEKKKLQAO 282
  |||
Db 273 NY--TJEMTIDSSMEDHNEINGSISNFNCC--PSSSEFDSK-----ESKNV--- 316
QY 283 YDLISYKQLEEAHNLISYLEKRIIDLTKKNENIKELDKINEIKNPPANGNTPTLLD 342
  |||
Db 317 ---INHTLRDKNNILITSSSEYIKDL-HNCFNFSNSDK-----NDSF 355
QY 343 KKKIIEHEKEIKETIAKTF--NIDSLFTDPLELEYLREKKNKIDISAKVTKESTE 399
  |||
Db 356 KLYGICEYDKYLDILEENASFFHNNVD-----EYGYDYNNKNTNLSNKKIQNNNE 408
QY 400 PNEVPNCVTPYPLSYNDINMLNEL-----NSFGDLINFDYTKEPSKIYTD-NERK 450
  |||
Db 409 NKK-----NNKNNNNNEVDYIKKDEDNVNSKVFYSOYNNNANONNEHTPEPLNN 457
QY 451 KFIENEIKERK-----IEKKKIESDKSYEDRSKSLNDITKYEKULMEIYDSKFRNN 503
  |||
Db 458 DYSYTIKKKKMNEEFMLVNNKRYVDHKE-----IIVYVCGATKS 498
QY 504 IDLTNFEKMGKRYSYVEKLTTHNTFASYE-----NSKNLEKLTALKYMEYSLRNI 558
  |||
Db 499 IIKMLIRHGMDLPLYIIVPYNNFHIDYAVLLSNGPGDKCDPLIKMLKSLTRKK 558
QY 559 VV-----EKELKYNNKLSIKIENELJETLVENIKKDEQ-----LFEKK 596
  |||
Db 559 IIFPICIGNOLGSLGCDYTKMYGNR--GVNOPYIOLVDNCTITSQNHGICLKKS 615
QY 597 ITKDE-----NKPDEKILEVSD-----YKVQYQKVLNKKIDELKKTOLILANVELIK 645
  |||
Db 616 ILKRELALSYINANDKSIEGISHKNGRFYSVQHP--EGNNGPBDTSPLFRKNFL-- 669
QY 646 NIHPNSYKQENKOEPPYLLYLKKEI----- 671
  |||
Db 670 --DIFNKKQYREYLYGIYIYIKKRVLLSGGLGICGAGEFDYSQTQAIKSLKEGITYV 727

```



```

QY 672 -----DKLVPMK-----VSLINEKKN-ITKESGDSNESPTEGE 708
DB 728 ILVNPINATVQTSKGLADKV-YFLPVNCEFEVETIKKEKPDFILCTFG----- 774
QY 709 INGQATTKG-----QOAGSALGSDSVQAO-----AOQOQAQPPVPVPPE 750
DB 775 --GQALNCAALMDOKKVLKKNCCOLGTSLESIRITENRPLFAEKLKEINERI-APYGS 831
QY 751 AKAQVPTPPAPVNNKTEVNSKLDY--LEKLYEF--LNTSYI-----CHKYILVSH 796
DB 832 AK-----NNQOAIIDANKIGYPLVYRTTSLGSLNSSFINEBELLEKCKIFL---- 880
QY 797 STWNE-----KILK-----QYKITEES-----KLSGCDPLDLFNIONNIPVMSWFD 841
DB 881 QJDNLEFIDKSLGKMEIEYELLROKKNCAIACNMENIDPL-----GIHTGDSIVVAPSQ 936
QY 842 SLNNSLSOLFMEYEKEMV-----CNL-YKLKNDKIKNLEAKKYSTSVKTLSSSS 893
DB 937 TISNTEYTFREIATKVIITHLNIIGECNIOFGINPOTGEYCIEVVARLSRSALASKAT 996
QY 894 MOPL-----SLTPQDK-----PEVSANDDTSHSTNLN 921
DB 997 GYPLATYISAKIALGYDILSLKNSITKKTACPEPSLDYITTKIPMDLKKFEFASMTMS 1056
QY 922 SLKLEENILSGKNKNTYQELLGQKS-----SENFY-----EKIL-----K 957
DB 1057 SKMSVGEVVSIGRT--FEESI-QKSLRCIDNYLGFNSYTCIDMDERKIIIEELKNPSRK 1112
QY 958 DSDTFNESEFTNPFVSKAD-----INSLNDESKRKKLEE--DINKLKK 999
DB 1113 RIDAIHOAHLNMPDKIHELHIDYWFLAKFYNIYNLOKRLKTLKEQLSPFDLKYPRK 1172
QY 1000 -----TLQISFDLYN-----KYKLEKLERLFDKRTYGVKKMOI 1032
DB 1173 HGFSPDKQIAHYLSFTSDNNNNNNNISCRVTENDVMKYREKL-GLEPHIKYIDLTSAEF 1231
QY 1033 KKLTL-----LLKQLESKLNSLNPKHYLONFSYFENKKK-----EAEI 1071
DB 1232 PALTYNLYLYTGOQEHVPLPLMKRRKKTCTLNKRNA-----NKKVHYVKNHLYNEV 1283
QY 1072 ATEMENT-----LENTKILKHYKGLVYNGESPPLKTLSESIOTEDN 1115
DB 1284 VDDKDTQLHKEKNNNNNMNSGAVENKCKLKESTYG-----YNNSSNCINT--NNIENN 1336
QY 1116 ---YASLENFVY-LSKLEGLKLDNLLEKKLSTYLS--SGHLIAELKEVI----- 1161
DB 1337 ICHDISINKIKVYTIINSNSISNNEVNE-TNLNCYSEBAGSHIHYGKEKESIGSDDTNI 1395
QY 1162 ---KKNKYTGNSPSEN--NPDVNNALLESYKFLPEGTVDATVSESGDILEOSQPRK 1214
DB 1396 LSAONSNNSNENMKNANVDVNLNDTKK--REDINTTTFPME--QONSVINNNKKE 1452
QY 1215 PASTHGAESNTITTSQAVNDEVDVYIIVPFGSESEEDYDDGOVVTGEVAPPSVIDNL 1274
DB 1453 NSSLKGDDEEDIVMNLKKNENYSYI-----NNVDCRKKMDGK-----NINDECK 1499
QY 1275 SKIENEYEVLYLPLAGVYRSKOLENNVM-----TFNVAVKIDILNSR-----F 1319
DB 1500 TYKKNKXKDM-----GLANNITVDELSTNGTSHSTNDHLYLDNNTSDELEGNKKNDMDYL 1553
QY 1320 NKRENFKNVLESDLPYKOLUTSNVYVKKPYFLNKEKDKLSSYNYIKDSIDTIDIN-- 1377
DB 1554 SKKESISN-----KNPGNSYVVDVYNNNEYKINKMKELIDENINDEYNNVNMN 1604
QY 1378 -----FANDVLGYIILSEKYSKSDLSIKKY--INDOGENEKKLPLNLNITETLX 1425
DB 1605 CSNYNNASAFVNGKDRNDLNDCELEKNDHYTKHYNRNNRSTTERMLAVNN----- 1659
QY 1426 KTYNDKIDFVYIHLLEAKVLYNT--YEKSNVYEVYKIKELNLTAKTJODLAFCKNNNVGIA 1483
DB 1660 -----EKESNHEKGHRNGLNKKKKNKEKNNEKN--KGNKND--KKNHYIYV-N 1701
QY 1484 DLSTDYNNHNLTKF 1498

```

```

DB 1702 KRNEYNNSNIESKF 1716
: :| | :| |
RESULT 38
US-08-755-587-44
; Sequence 44, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; type: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-755-587-44

Query Match 3.5%; Score 292.5; DB 3; Length 3418;
Best Local Similarity 18.6%; Pred. No. 3.3e-07;
Matches 382; Conservative 288; Mismatches 717; Indels 663; Gaps 92;
QY 7 LCSFILE-----FIINTQCVTH--ESYQELVKLEALEDAVLGTGSLPQKEKMYLNECTSGT 60
DB 160 VCGSLFHTPKFYKGRQTPRHISELSIGAEDVDPDMSWSSSLATPTLISSTVLIYVNEASET 219
QY 61 AVTTSTPGSKGSVASGSGGSVASGGSVAS--GGSVASGGSVASG--SGNSRRTNNS 114
DB 220 VEPHDTTAVVKSYSFNHDESLKKNDRFIASVIDSENTNOREASHGFGKTSQNSFKVN-- 277
QY 115 DMSSDDAKSYADLKRVRNNYLLTIKELKYPQLFDLTNH-MULTIDNHGFRYL--IDGY 171
DB 278 -SCKDIGHGSMPIV-----LEDEVYETVVDTSSEDSFSLCSKCRTKNLOKVPTS 326
QY 172 EEINLELYKINFEDLLRAKLVNDCANDYCOI-----PFMLKIRANEDVLKLVFGRKP 227
DB 327 KTRKKTFHFAN-----ADECEKSKNQYKEKSFYSVEVPNDPTDLDSNV-AHQRP 375
QY 228 LDNIKDNVGMEDYIKKNNKTIEINELIEESKKTIDKKNKATKEBEKKLYQAOYDLSI 287

```



```

Db 376 FEESGSDKISK-----EYVPSLACEMSQÜLTSLGSLNQAQMEKIPLLHISSODONI 423
Qy 288 YNKÖLEEHANLISVLEKRIIDLTKKENENIKELLDKINET-KNPPRANGTPTMTLLDKNK 346
Db 424 SEKDLDPTEN-----KRRKDFLTSEN--SLPRISSLPSEKPLNEETVANK-DEEÖH 473
Qy 347 IEEH-----EKEKEIAKTIFKIFIDSIF--TDPLELEY 378
Db 474 LESHTDCLLAKQAISGSPAASSFOGIKISIFRIRESKEFTFNASFSGHMTDP-NFKKE 532
Qy 379 LREKKNIDISAKVETKE-STEPNEYPNGVYPLSYNDINNALNELNSEGLINPEDYRK 437
Db 533 TEASGSEIHTVCSQKEDSLCPNLIDNG-SMPATTQNSVALKNAAGLSTL----- 583
Qy 438 EPSKNIYDNEKKKFINIEKEKIKIEKKIESDKKSYEDRSKSLNDITKEYEKLNEYD 497
Db 584 -----KKTINKFIYAIHDETFYKGRKKIPKÖKS-----ELIN-CS 617
Qy 498 SKFNNDIDLTFNEKMMGRYSYKVEKLJHNTFAEYEN-----SKHNLEKLTFA 546
Db 618 AOFENNA-----FEAPL-----TFANADSGLHSSVKRSCSQNDSEPTLS 658
Qy 547 LKYMEDYSLR-----NIVEKELEKYKNLISKIENEIETLVENIKKDEOLFERR 596
Db 659 LTSSFGTILKRCRNETCSNMNTVISQDLDYKRAKCNK-----EKLÖFTTPPADSLSCLO 713
Qy 597 ITKDNKDKDEKILEYSDIVK-----VOYQVYLANNKIDELKKTÖLILKNEVKHNI 647
Db 714 EGQCNDKRSK--KYSDIKEEVYLAACHPVQHSKVEYSDTPDQOSQSL----- 760
Qy 648 HVPNSYKOENKOEYLYLVLEKKEIDKLKVFMPKVESLINEEKNKKTGEOQNDSEPTSG 707
Db 761 -----YDHENAST--LITPTSKVYLSNV-----MISRGESYKMSDKLKGNYESV 807
Qy 708 EITGOATTPQGOAGSALEGSVQAQAOEQAPPVYVPEAKAQAQVTPPAPVY-NKT 766
Db 808 ELTKRIPMEKNO-----DYCALNENYKN-----VELLPREKYMRVASFPRKVOFQON 854
Qy 767 ENVSKLDVLEKLYEFLNYSYICHKIIVSHSTMNEKILKOYITTEBESKLSGCPDL 826
Db 855 TN--LRVIOKMOE--ETTJIS-----KITVNPSE-----ELF 883
Qy 827 FNIONNIVMYSMPDSLNNLSQLPMEIYEKEMVNLKLDKNDKIKNLLEBAKRYSTV 886
Db 884 SDNENNF--VFQVANNERNNALGNTELHEDTLTCVNEPIFNKSTMVLYGDIQDAIY 941
Qy 887 KTLSSSSMOPLSLTPÖDKPEVASNDTSHSTNLNLSKLFENIISLGKKNITYOE---L 942
Db 942 ---SIKRDLYVYVLAENKNSVKQHIKMTIGODLKSDISL--NIDKIPKNNNDYMKMAGL 996
Qy 943 IGOKSESNF-----YEKILKDSPTFYNESFTNFKSKA--DOJNSLN-DES 985
Db 997 LGPISNHSFSGSFRTASKKEIKLSHNTIKSKMFKDIEBOYPTSLACVEIVNTIALDNO 1056
Qy 986 KKKLEEDINKLKTLOLSF--DLYNKYKLEERLPDK-----KTVGKYKMOIKKLT 1036
Db 1057 KKLKPOQINIVYSAHLÖSSVYVSDCKNSH-ITPQMLFSGKÖDNSMHNLLPSKAITTELS 1115
Qy 1037 LKKEÖLES--KLNSLNPKHVLQ--NFSVFENK-----KEAEI----- 1071
Db 1116 TILBESGQOFETQFRKPSYILOKSTFEVPEQOMTILKTTSECHDADHLVIMNAPISQ 1175
Qy 1072 AETENTLENTLILKHYGVK-----YNGESSPLKLTSEESI 1110
Db 1176 VDSKQFEGTVEIKKFKAGLLKNDCKNSASGYLIDENEVEGFYSAGTKL-NVSTEL 1234
Qy 1111 QTEDNYASLENFYVLSLEGLKLDNLNEKKKLSYSSGLHLLIAELKVIKKNYVTSNS 1170
Db 1235 Q-----KAVKLFSDE-NISEETSAVHPHISLSSSKCHDSVSM--FIEHNDKT 1282
Qy 1171 PSENN-----TDVNNALSEYKFLPECTDVATVVSSEGSOTLE---QS 1210

```

```

Db 1283 VSEKNNKCOLLIÖNNIEMTGTFFVEIETENYKR-NTENEDNKYTAASRSHNLEFGSDS 1341
Qy 1211 ÖPKKRASTHVGAESNTITTSQVNDDEVYIIVPIFGESEEDYDL-----G 1257
Db 1342 SKNDYVCIHKDETDLLFTFOHNIKLSQGFKKEGNTQIKEDSLDLEFVAKAÖEACHG 1401
Qy 1258 QVVTGEAVTPSYIDNLSIKIENEYEVLYLKLAGVYRSKLOLENNVMF-----N 1308
Db 1402 MTSNKEQALATTEONIKDEF--SDTFPÖTASGNISYAKELFNKIVFPOCKPEELHN 1459
Qy 1309 VVVKDILNSRFKRENFKNVL--ESDLIPYKDLTSS-----NYVKDPYKFLNKRKD 1359
Db 1460 FSLNSELHS--DIRKKNKMDLSYEEDIVKHKILKESVYVGQNOVY-----TFQGOPEKD 1513
Qy 1360 KFLSSYNTIKDSIDPDINFANDVLYKYLSEKY--KSDLSIKKYINDK-ÖGENE--- 1412
Db 1514 E-----KIKED-----TILGFHFASGRKVIKAKESLDKVNLPEDKEÖGTSSEITS 1558
Qy 1413 -----KYLPLNIEETLKTY-----NDK-----IDLFY----- 1436
Db 1559 FSHQNAKTLKREACKDLELACETIEITPAAPCKEMÖNSLNDKNLVSIETVVPKRLSD 1618
Qy 1437 -----IH-----NDK-----IDLFY----- 1438
Db 1619 NLCRÖTENLKTSKSIPLKVKVHENVKEKTAQSPATCYTNQSPYSIENSALAFYTSCKRK 1678
Qy 1439 -----LEAK-----VLATYTB-KSNVEYKIKELWYLTQDK 1469
Db 1679 TSVQSÖTSLLEAKKWLREGIFDQPERINTADVGVNLYENNSNSTIENDKXHLSEKODT 1738
Qy 1470 -LADFRKNNFYGIDLDSTDYNNHNLTKF-FLSTGM--VFENLA-----KTVSLNLD 1518
Db 1739 YLSNMSMSYSYHSD--EYVDSRLSKNLKIDSGIEPLKAVEDOKNTSEKVISNVVD 1796
Qy 1519 GN-----LOGMLNISOH 1530
Db 1797 ANAYPQTVNEDICVEELVYTSSPCKKNKNAIKLISNSNNEFVGPAPRIFASGKIRLCSH 1856
Qy 1531 ÖCVKK-----ÖCPÖNSGCPFHLDRECKKLLWYKÖGDKC 1566
Db 1857 EYIKRVKIDDSKYVIFTKENNEKSKICÖTKIMAGCYEALDSDSD--TLNNSLNDNJC 1913
Qy 1567 VENPNPTC-----NENNGCDADAKCTEEDSGSNGKKTICECT-----KPSYP 1610
Db 1914 SMHSKIVFADIOSEELIÖHÖNMÖSGLEKYSKISPODVSLTSDI-CKCSIGLHKSYSVA 1972
Qy 1611 LFDGIFCSSS 1620
Db 1973 NTCGIFSTAS 1982

```

```

RESULT 39
US-08-603-753D-4
; Sequence 4, Application US/08603753D
; Patent No. 5891857
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARLES A. TAYLOR, JR.
; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
; CITY: DURHAM

```



```

STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603/753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HS043746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION: regulatory effect on growth of human mammary cells.
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4

Query Match 3.4%; Score 288.5; DB 2; Length 3418;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 382; Conservative 287; Mismatches 718; Indels 663; Gaps 92;

```

```

QY 172 EEINELLKLNFEYDILLRAKLANDVCANDYQI---PENIKIRANELDLKIVFGYRKP 227
DB 327 KTRKKIIFHEAN-----ADECEKSKNQVKEKSYFSEVBPNDPLDSNV-AHQKP 375
QY 228 LDNINKDVGKMEYIKKKNKTININELIEESKTTIDKNNNAKKEEKKLYQAQYDLSTI 287
DB 376 FEESGDSKISK-----EVPPLACEMSQLTSLGNGAQMERIPPLHLSISCDQNTI 423
QY 288 YNQLLEAHNLISVLEKRIIDTLKKNENIKELDKINEI-KNPPANGNTPTLIDKNNK 346
DB 424 SEKDLDLTEN-----KRRKDFLTSEN--SLRISSLPKSEKPLNDETVYNN-RDEQH 473
QY 347 IEEH-----EKEIKEIAKTIKFNISLF---TDPLEEY 378
DB 474 LESHTDCLIAVQAISGTSVAVASSFOGIKKSIPRIRHSPKPEFYNASGCHMDP-NFKKE 532
QY 379 LREKKNIDISAKVEIK-EETPEYENGVTYPLSYNDINNALNELNSFDLDLNPDTYK 457
DB 533 TEASESGLEIHTVCSQEDSLCPNLIDNG-SWPATYTONSVALKMAGLISLTL----- 583
QY 438 EPSKNITDNERKKFNEIKIEKIKIEKKIESDKSYEDRSKSLNDITKEYEKLNEYD 497
DB 584 -----KKTNKFYAIHDETFYKCKIIPKOKS-----ELIN--CS 617
QY 498 SKFNNNIDLTNFEKMMGKRSYVEKLTHTNTPASYEN-----SKHLEKILTKA 546
DB 618 AOFEPANA-----FPAPL-----TFANADSGLLHSSVKRSQNSSEPTLS 658
QY 547 LKYMEDYSLR-----NIVYEKELKYKKULSKIEINEIETIVNKKIDDEQLFEKK 596
DB 659 LTSSEFTILKRCGRNETCSNNVTIISODLYKEKCNK-----EKLOLFTPEADSSCLQ 713
QY 597 ITRDENKPDKEILEVSDIVK-----VOQKVLMLNKIDELKTLQILKLVKLNHI 647
DB 714 EGQENDPKSK--KVSIIKEVLAACHPYOHSHKVEISDDIPQOSKSL----- 760
QY 648 HVDNSYKQENKQEPYLYIVLKKETDKLVMPVYESLINEKKNKITEGOSDSEPTSEG 707
DB 761 -----YDHENAST--LILPITSKDYLSNLY-----MISGRKESYKMSDKLKNNEESDV 807
QY 708 EITGQATTRPGQAGSLBEDSVQAQAOEOKOAPPVYVVPVEKKOVPYPPAPVNN-NKT 766
DB 808 ELTKNIPMEKNQ-----DVCALNENYNN--VELLPPEKYMVASPSRRKQVON 854
QY 767 ENYSKLDYLEKILEYFLNTSYICHKYLIVSHSTNEMKILKOYKITEEESKLSSCDPLDL 826
DB 855 TN--LRVIQKNOE--ETTSIS-----KITVNDSE-----ELF 883
QY 827 FNIONNIPVMYSMFDLSLNNLSQLFMEIYKEMVCNLKTKNDKIKNLLEAKKYSTSV 886
DB 884 SDNENNF--VFQVANERNNLALGNTKELHETDLTCVNEPIFKNSTWLYGDTGDKATQV 941
QY 887 KTLSSSSMOPLSLTPQDKPEVSANDTSHSTNINLSLKLFEINLSLCKKNITVOE--L 942
DB 942 ---SIKQDLYVLAEEKNSVKOHIKMTLQODLKSQISL--NDKIPKQNDYNNKRAKL 996
QY 943 IGOKSSSENF-----YEKILKSDFTFYNESFTNFKSKA--DDINSLN-DES 985
DB 997 LGRISNHSFGSFRITASNKEIKLSEHNKSKMFKNIEQYPTSLACVEIYVTLALDNG 1056
QY 986 KKKLEEDINKLKTLOLSF--DLYNKYKLEKLELPDK-----KTVKRYMQIKKL 1036
DB 1057 KKLKSPQISINTVSAHLQSSVAVSDCKNSH-ITPOMLFSEKODFNSNHLTPSQAIEIYELS 1115
QY 1037 LKEQOLES--KLNSLNPKVILQ--NFSYFNNK-----KEAEI----- 1071
DB 1116 TILLESQSOEFTQFRPSYLIQKSTFEVENDQTIKTTSECRDLDLHVINNAPSIGQ 1175
QY 1072 AETENTLENTKILKLBKKGIVK-----YNNSESSPLKTLSESI 1110
DB 1176 VDSKQEPGEIVETKRRFAGLKLKNDCKNSASGIYLTDENVGFRGFIYSAHGKIL-NVSTFAL 1234

```



```
QY 1411 QTEVDYASLENEFKVLSKLEGLKDNLEKKKLSYSSGLHLIAELKEVIRKNKNTGNS 1170
D 1235 Q-----KAVKLFSDIE-NISETSABVHPISLSSSKCHDSVSM---KRIENHNDKT 1282
QY 1171 PSENN-----TDVNNALASYKKFLPEGTDAVAVVSESGDPLE-----OS 1210
D 1283 VSEKNNKCOLILONNIEMTGTGTFVEITEYENKR-NTENEDNKYTAASRNSHLEFDGSDS 1341
QY 1211 QPKRPASTHVGAESNTTTSQNVDEVDVLIIVPIFGESEEDYDL-----G 1257
D 1342 SKNDTVCIHKDETDLLFDQHNICKLSGQFMKEGNTIKEDLSDLTFLEVAKAQEAACHG 1401
QY 1258 QVATGEAVTPSVIDNILSKINEVEVLKPLAGYVRSKLKOLENNVMTF-----N 1308
D 1402 NTSKKEQULTAKTEONIDFET--SDTFPQASGKNISVAKELFKIYNFPQKPEELHN 1459
QY 1309 VNVKDILNSRFNRKRENFNVL--ESDLPYKDLTSS-----NVYKDPYKFLNKKERD 1359
D 1460 FSLNSELHS--DIRKNKMDILSYETEDIVKHKILKESVPGVGNOLV---TFQOQPERD 1513
QY 1360 KFLSYNYIKDSIDTIDIFADVIGYKILSEKY---KSDDSIKKYINDK-OGENE--- 1412
D 1514 E-----KIKEP-----TLGEFHTASGKKVIAKESLDKYNLFDEKRGTSSETTS 1558
QY 1413 -----KYLPLNINIEETLYKV-----NDK-----IDLFV----- 1436
D 1559 FSHQWAKTKYREACKDLELACETIETAPACKCKEMQNSLNNDKMLVSEIETVYPPKLLSD 1618
QY 1437 -----IH----- 1438
D 1619 NLCROTENLKTSKSIFLKVKYHENEKETAKSPATCYTNOSPYSVIENSALAFYTSCKSRK 1678
QY 1439 -----LEAK-----VINYTYE-KSNVEYKIKELNYLKTIDDK 1469
D 1679 TSVSOTSLLEKAKKMLREGIFPGQPERINFADYGVNYLYENNSNSTIAENDKNHLSKQDT 1738
QY 1470 -LADFERNKNNFVGIADLSTDYNNHNLTK-FLSTGM--VFENLA-----KTVLSNLLD 1518
D 1729 YLSNMSMSNSYSYHSD--EYVNDSGYLSKNKLDGIEPLVKVNEQOKTSSKVISNKK 1796
QY 1519 GN----- 1530
D 1797 ANAYPQTVNEDICVEELVTSSPCKNKNNAIKLISNSNNEFVGPAPRIASGKIRLCSH 1856
QY 1531 QCYVK-----OCPQNSGCFRHLDERECKCLANYKQEGDKC 1566
D 1857 ETIKKVKDIFTDSKTYKKEKNNENKSKICQTKIMAGCTEALDDSD---ILHNSLDNDEC 1913
QY 1567 VENPPTC-----NENNGCGADADAKCTEEDSGSNGKITCYECT---KPDSP 1610
D 1914 SMHSKVPADIOSFEIILQHNQNMGSLEKVISKISPCDVSLSTSDI-CKCSIGKILHSVSA 1972
QY 1611 LFDGIFCSS 1620
D 1973 NTCGIFSTAS 1982

RESULT 40
US-09-099-753-4
; Sequence 4, Application US/09099753
; Patent No. 6149903
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
```

```
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
STREET: BOULEVARD
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/099,753
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-09-099-753-4
Query Match 3.4%; Score 288.5; DB 4; Length 3418;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 382; Conservative 287; Mismatches 718; Indels 663; Gaps 92;
QY 7 LCSLF-----FTINTOCVTH--ESTYQELVKRLALEDAVLTGYSLFOKEKKVNLGSGT 60
D 160 VCGSLFHTPKFVKRGQTPKHISESIGAEVDPDMSSSLATPPTLSVTLIVRNEASET 219
```



```

OY 61 AVTTPGSKGSAVSGSGSVASG--GSGVAGSGVAGSG--SGNSRRTNPS 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 VEPHDTAVKVSFSNHDSEIKKNDKRFIASVDTSEMTNOREASHGFGKTSGMSFVN-- 277
OY 115 DNSSODSAVADLKHVRNRYLLTJELKYPOLFDLTNH-MLTCDNIHGEKYL--IDGY 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 -SCKDHGKSMRVN-----LEDEYETVADTSEEDSFSLCFSCKPTKMLQWRTS 326
OY 172 EENELLYKLNFFDLBRKLNDCANDYCOI-----PENKIRANEDLYKRIYFGRKP 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 KTRKKLTFHEAN-----ADECKSKNOVAKESYSEVEPEPNDTDLDSNV-AHQKP 375
OY 228 LDNIKNVGMEDYIKRNNKTENINELIEESKRTIDKNKNAKEBEKKLYOAOYDLST 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 FEGSGDKISK-----EVPPLACEMQSLTSLGNGAQMKRIPLHISCDQNI 423
OY 288 YNKQLEAHNLISVLEKRIDTLKKNENIKELDKINEI-KNPPRANSNTPTNLLDKNKK 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 SEKDLDTEN-----KRRKDLTSEN--SLPRISLPSKSEKPLNEETVVK-RDEEQH 473
OY 347 IEKH-----EKEIKEIKTIKFNIDSLF---TDPLELEY 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 LBSHTOCLIAVKAISGTSFVASSFOGKIKSIRRESPEETNASTSGHMDP-NFKKE 532
OY 379 LREKKNIDISAKVETRE-STEPNEYNGVTPPLSYNDINNALNELNSFGLINPEDYTK 437
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 533 TEASESGLEIHVYCSQKEDSLCPNLIDNG-SWPATTTQNSVALKNGLISTL----- 583
OY 438 EPSKNITTDNEKFKTNEIKETIKIKKIKIESKSYEDRSKSLNDITKEKLLNIYD 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 584 -----KKTNNKFTYAIHDEFYFGKKIPKOOKS-----ELIN-CS 617
OY 498 SKENNNIDLTNFEKMMGRKRSYKVEKLTJHNTFASYN-----SKHMLEKTKRA 546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 618 AOEPAAN-----FEAPL-----TTANADSGLLHSVSKSGQNDSEPTLS 638
OY 547 LKYMEDYSLR-----NIVERELKYNNLISKIENEIETLVENIKDEOELEKK 596
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 639 LITSFETILKCSRNFTCSNNTVISOOLDYKKEAKCNK-----EKLOLFIPEADSLCLO 713
OY 597 ITDEKAPPEKILEVSDIYK-----VOVOGVLMMNKIDELKKTQOLLKKNVELKNI 647
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 EGCEGENDPNSK-KVSDIKIEEVLAAACHPVQHSKVEYSDTDFOSQSL----- 760
OY 648 HVPNSYKOEKOBEPYLLIVLKEIKEDLKVMPKVESLINEEKNKITEGOSDNSESTEG 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 761 -----YDHENAST--LILPTSKDVLNLY-----MISGKESYKMSDKLCKNNYESD 807
OY 708 EITGQATTFRGOAGSALBGSVQAOAOEOKQAPVVPVPEAKAOVTPPAV-NKT 766
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 808 ELTKNIPMEKNQ-----DVCALNENYKN-----VELLPPEKYMVASPSRKYOFNQN 854
OY 767 ENVSKIDYLEKLEFINTSYICHKYLIVSHSTNNKIKQKITTEESKESLSCDPLDLL 826
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 855 TN--LRVJOKNOE-ETTSIS-----KITVNDSE-----ELF 883
OY 827 FNIOANNIPWYSMFDLSNLSOLFMEIYERKENCYLKLDNDKIKNLLEAKKYSTV 886
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 884 SONEHNF--VFOYANERNNALGNTKELHETDITLCVNEPIFKNSTWLVYDITDQKATQY 941
OY 887 KTLSSSMOPLSLTPODKPEVNSANDTSHSTNINNSIKLEFENILSLGKKNITQE--L 942
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 942 ---SIRKDLVYLAEBKNKSVKOHIKMTLQODLKSDISL--NIDKIPKNNDMYMKWAGL 996
OY 943 IGOKSENF-----YEKILKSDPFYNESFTNFVSKA--DDINSLN-DES 985
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 997 LGPISHSFEGSFRFASNKEIKLSEHNKSKMFEIDIEQYPTSLACVAVITVTLADNQ 1056
OY 986 KRRKLEEDINKLKTLOLSF--DIYNKYKLTLEPLDK-----KTGVKYMOKIKLT 1036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1057 KKLKSKQSINTVSAHLQSSVYVSDCKNSH-ITPOMLFKQDFNSNHLTSPQSAEITELS 1115

```

```

OY 1037 LKKEOLES--KLSNLNPKHVIQ--NFSVEFNK-----KEAEI----- 1071
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1116 TLEESGSOFEFTQFRRKPSYILOKSTFEVENDOMTILKTSEBCRADLHVMNAPISIQ 1175
OY 1072 AETENLENTKILKHKYGLV-----YNGESSPLKTLSEESI 1110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1176 VDSSKQEGEVEITKRRFAGLKNDCNKSASGYLTDENEGFRGFYSANGTKL-NVSTEAL 1234
OY 1111 QTDONVASELNFVLSKLECKLNDKLNLEKRLKSYLSGLHNLIAELKEVYKKNYTGNS 1170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1235 Q-----KAVLFSIDIE-NISEETSAEVHPISLSSKCHDSVSM--FKIENINDKT 1282
OY 1171 PSENN-----TDVNNALSEYKFLPEBDTAVTVSESGDYLE---OS 1210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1283 VSEKNNKCOLLIONNIEMTGTVEEETENYKR-NTENENKRYTASRNSHNLEFGSDS 1341
OY 1211 QPKKPASTHGAESNTITTSQNVNDEVDVYIYPIGSEBEDVDL-----G 1257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1342 SKNDYVCIHDEFTDLFTPDQHNICLKLSQGFMEKNGTQIKEDLSDLFLEVAKAQCACHG 1401
OY 1258 QVVTGEAVTPSVIDNILSKIENEYEVLYKPLAGVRSKOLENNVMTF-----N 1308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1402 NTSNKEQLTATKTEQNIKDEP--SDTFPOTASGKNISVAKELFNKIVNFPDQKPELHN 1459
OY 1309 VNVKDIINSFENKRENFKNL--ESDLPIYKDLTSS-----NYVVKDYPKFLNKEKD 1359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1460 FSLNSELHS--DIRKKNMDLSYEEDIVYKHKILKESVPVGTNOVL--TFQGGPERD 1513
OY 1360 KFLSSYNYIKDSIDTQINFRANDVLYGYKILSEKY--KSDLSIRKYYNDK-OGENE-- 1412
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1514 E-----KIREP-----TLGFHTASGKVKYAKESLDVYKKNLPDEKEGTSEITS 1558
OY 1413 -----KYLPFLNNIETLYKTV-----NDK-----IDLFY----- 1436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1559 FSHQWAKTLKRYRACDLELACETIETTAAPCKCKEMQNSLNDKNLYSIETVVPKLLSD 1618
OY 1437 -----IH----- 1438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1619 NLCROTENLTKSIFLKYVHVENKETAESPATCYTNOFPSYIENSALAFYTSCSRK 1678
OY 1439 -----LEAK-----VLNITYE-KSNVEYKIKELNYLKTQDK 1469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1679 TSVSOTSLEAKKWLREGITFDGQPERINTADYGNVLYENNSSTIAENKKNLSEKQDT 1738
OY 1470 -LADFKNNNFVGIADLSTYNNHNLTK-FLSTGM--VEENLA-----KTVLSNLLD 1518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1739 YLSNMSMSNGYSYHD--EYVNDGYLSKKNKLDGIEPVLKVNEDQKNTSFSKVISNVKD 1796
OY 1519 GN-----LOGMLNISOH 1530
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1797 ANAYVPOTVNEDICEBELVYSSPCKNKNAIKLISNSNNEFVGPAPRIASGIRLCSH 1856
OY 1531 QCVKK-----OCPONSGCFRLDREBECKLLANKQEGDK 1566
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1857 ETIKKYKADITDSFSKYIKENNEKSKITCOTKIMAGCYELADSED--ILHNSLNDDEC 1913
OY 1567 VENPNPTC-----NENNGCDADACTEEDSGSNGKITECT-----KPDSTP 1610
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1914 SMHSKWFADIQSEILJOHNQMSGLEKYSKISPCDVSIETSDI-CKCSIGLHKHSVSSA 1972
OY 1611 LFDGIFCSSS 1620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1973 NTCGIFSTAS 1982

```

```

RESULT 41
US-08-986-106-4
; Sequence 4, Application US/08986106
; Patent No. 6177410
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: KING, MARY-CLAIRE

```


APPLICANT: STEINER, MITCHELL S.
APPLICANT: ROBINSON-BENION, CHERYL L.
APPLICANT: THOMPSON, MARILYN E.
TITLE OF INVENTION: THERAPEUTIC METHODS FOR
TITLE OF INVENTION: PROSTATE CANCER
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,106
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,753
FILING DATE: 20 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
FEATURE:
NAME/KEY: BRCA2 protein
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granlin box domain at
RELEVANT RESIDUES IN SEQ ID NO: amino acids 3334-3344
US-08-986-106-4

Query Match 3.4%; Score 288.5; DB 4; Length 3418;
Best Local Similarity 18.6%; Pred. No. 5.4e-07;
Matches 382; Conservative 287; Mismatches 718; Indels 663; Gaps 92;

QY 7 LCSFLP-----FIINTQCTH--ESYQELVKLELEDAVLGYSLPQEKVYLVNGTSGT 60
DB 160 VCGSLFHTPKFVGKQRTPKHISESGAEVDPDMSSSSSLATPPPTLSVTLVYRNEASEET 219
QY 61 AVTSTPSKSGVASGSGSVASGVSAS--GGSVASGSGVASGVSAS--SGNSRRTMPS 114
DB 220 VPHDITANVSYSFNSHESLKKNDRTASTYDSENTQREASHGFEKTSQNSKFN-- 277
QY 115 DNSSDSAKSYADLKHRYRNYLLTIKELKYPQLFDLTNH-MLTLDNTHGEKYL--IDGY 171
DB 278 -SKDHICKSPNV-----LEDEYETVVDTSSEDSFSLCFSKCRKNNLQKVRTS 326
QY 172 EFINELLYKLNFYDLAKLNDVCANDYCOI---PNNLKIRANELLVYKLVGYGKRP 227
DB 327 KTRKKIFHEAN-----ADCEKSKNOYKEKYSEFSEVEPNDTDLPSNV-AHOKP 375

QY 228 LDNKDNVKNMEDYIKKNNKTJENINELLEESKRTIDKNNKATKEEKKKLYQAOYDLST 287
DB 376 FESSGDKISK-----EYVPSLACESSQLTSLGNAQOKKPIPLHISSCDQNI 423
QY 288 YNKOLEEAHNLISLEKRIIDLTKKNNIKELDKINEI-KNPPANSNGTPTLLDKNK 346
DB 424 SEKDLDTEN-----KKKDDLTSEN---SLPRISSLPSKSKPIINEETVYVK-RDEQ 473
QY 347 IEER-----EKEKELAKTIKFNDLDF---TDPLELEY 378
DB 474 LESHDCILAAVKAQALISGTPVASSPQGIKSLFRIRSEPKETFNASFQHMIDP-NFKKE 532
QY 379 LREKKNIDISAKVEYKE-STEPNEYPNGVYPLSYNDINNALNELNSFGDLINPDYTK 437
DB 533 TEASESGLEIHYVCSQKEDSLCPNLIDG-SMPATTONSVALKNAAGLSTL----- 583
QY 438 EFSKNIYDNEKKRKPIINEIKKIKIEKKIESDKSYEDRSKSLNDITKEYEKLNEYD 497
DB 584 -----KKTKNKFIVAHDETFYKGRKIPQOKS-----ELIN-CS 617
QY 498 SKFNNDITLNEPKMMGRKYSKYVEKLTHNHFAYEN-----SKNNLEKTKA 546
DB 618 AOEFAANA---FEAPL-----TFANADGGLHSSVKSCSQSDNSEPTLS 658
QY 547 LKYMEDYSLR-----NIVEKEELKYXNLSIKIENEIETLVENIKKDEOLFERR 596
DB 659 LTSSFGTILKRCSSRNETSNNNTVISQDDLYKRAKCNK-----EKQLPTTPBADSLSCQ 713
QY 597 ITKDKNKPKDEKILEVSDIVK-----VOYQKVLNNKIDELKKTQILIKNVELKINI 647
DB 714 EQCENDPKRSK--KVSIDIKIEVILAAACHPVQSHKVEYSDTQFOSOKSLD----- 760
QY 648 HVPNSYKQENQEPYLLVYKLEIKDKLVMPKVSLSLNEEKNKTKTGOSDNSPRTG 707
DB 761 -----YDHEMNST---LITPTSKDYLSNLY-----MISRGSEYKMSDKLGNAYESDV 807
QY 708 EITGQATTKPGQAGSALEGSVQAOAEOBOKQAPVVPVPEAKAQVPTPAPVY-NKT 766
DB 808 ELTKNIIPMEKNQ-----DVCALNENYKN---VELLPREKMYRAVSPRKYQAFON 854
QY 767 ENVSKLDYLEKLYEFLNTSYICHKYLIVSHSTANEKILKYKITYKEESKLSGCCPLDLL 826
DB 855 TN---LRVIOKQOE--ETTSIS-----KITVPDPSE-----ELF 883
QY 827 FNIONNIIVMYSMPDLSNLSQLFMETIEKEMVONLKKIKNDKIKNLEBAKKVYSV 886
DB 884 SDNENN--VFQVANERNNLALGNTKELHETDLCTVNEPIFNKSTWVLXGDTGQATQV 941
QY 887 KTLSSSMQPLSLTPQDRPEVSANDTSHSTLNNSLKLFENILSLGNKNITYOE---L 942
DB 942 ---SIKKDLVYVLAEEKNSYKQHIKMTLGQDLKADISL--NIDKIPPEKNNDYMKKAGL 996
QY 943 IGOKSSENF-----YEKILKSDTFYNESFTNFVSKA--DDINSIN-DES 985
DB 997 LGPISNHSFGGSFRTASNKEIKILSEHNIKKSMFKEIDIEQYPTSLACVEIYNTIALNQ 1056
QY 986 KRKKLEEDINKLKTLOLSF---DLXNKYKLELELPKK-----KVVGKYMOKIKLT 1036
DB 1057 KKLSPQISINIVYSAHLQSSVYVSDCKNSH-ITPQMLFSKODPNSNHNLTPOKAEITELS 1115
QY 1037 LLKEOLES--KLNSLNNKHYLQ--NFSVFFPKK-----KEAEI----- 1071
DB 1116 TILESSEGFETQPRKRSYILQKSTFEVPEKQMIILKTTSECCDADLHYMNAPIGQ 1175
QY 1072 AETENTLENTKILKHYGLVK-----YNGESSPLKTLSEST 1110
DB 1176 VDSKQFEGTVEIKRKFAGLLKNDCKNSGASGYLTEDENEGFRGFSAGHCTL-NVSTAL 1234
QY 1111 QTEDNAYSLNPKVSLKLEGLKQNLNLEKKKSLTSLSGHLHIALEKVIKNNKYTNS 1170
DB 1235 Q-----KAVKLFSDIE-NISETSAEVPHISLSSKCHDSVSM--FKIENHNDKT 1282
QY 1171 PSENN-----TDVNNALSEYKKFLPECTDAVAVVSESGSDLE---QS 1210


```

Db 1283 VSEKNNKCOLILONNIEMTGTGFEETENYKR-NTENEDNKYTASRSHNLEFDGSDS 1341
Qy 1211 QPKKPASTHVAGASNTTTSQNVDDVDYIYPIGESEEDDL-----G 1257
Db 1342 SKNDVCIHDEDTLFTDQHNCLKSGQFMKEGNTQIKLEDSLDFEVARAQAACHG 1401
Qy 1258 QVVTGEAVPVSVIDNLISKIENEYEVLYKPLAGVYRSKLENNVMTF-----N 1308
Db 1402 NTSNKKQLATKTEQNIKDEET---SPTEFQTASGKNISVAKELFNKIVNFDQKPEELH 1459
Qy 1309 VNVKDLNFRKRNKFNKVL---ESDLPIKDLTSS-----NYVVKDPYKPLNKKRD 1359
Db 1460 FSLNSELHS--DIRKNKMDLSEETDIYKHKILKESVPVGTGNOLV---TEQOGPERD 1513
Qy 1360 KFLSSTNYIKSIDTDINFDVANDIAGYKILSEKY---KSLDLSIKKTIINDK-OGENE-- 1412
Db 1514 E-----KIREP-----TILGFHTASGKVKAKESLDKVKNLDEKEGOTSEITS 1358
Qy 1413 -----KYLPLNNIETLYKTV-----NDK-----IDLTV----- 1436
Db 1559 FSHQWAKTLKREACKDLDELACETIETAPKCKEMQNSLNDKNLVSITVVPKELSD 1618
Qy 1437 -----IH----- 1438
Db 1619 NLCROTENIKTSKIFLKVKYHENVETAKSPATCYTNOSPYSVIENSALAFYTSCSRK 1678
Qy 1439 -----LEAK-----VLNPTYE-KSNVEYKIKELNLTQIDK 1469
Db 1679 TSVSQTSLFAKKWLRGIFDGOPERINTADYGNITYENKNSSTAENKHNHSEKQDT 1738
Qy 1470 -LADFKNNNFVGIADLSDYNNHNLTK-FLSTGM--VEENLA-----KTVLSMLD 1518
Db 1739 YLSNSSMSNSYSYHSD--EYVNDGYSLSKKNKLDGIEPLVKVNEQDKNTSFSKVISWVK 1796
Qy 1519 GN----- 1530
Db 1797 ANAYPOTVNEDICEVELYVSSPCKNKNAIKISISNSNNEVGPAPRIASGKIRICSH 1856
Qy 1531 QCYK-----OCPONSGCFRHLDERECCLLNKOEGDKC 1566
Db 1857 ETIKKVKDIFTDSFKYIKENNENKSKICQTKIMAGCYEALDSDSD--ILHNSLDNDEC 1913
Qy 1567 VEPNPTC-----NENNGCDADAKCTEEDSGSNGKITCECT-----KPDSTP 1610
Db 1914 SMHSVYFADIQSEELIQHNQMSGLEKYSKISPCDVSLETSDI-CKCSIGKLHKSVSSA 1972
Qy 1611 LFDGIFCSS 1620
Db 1973 NTGIFSTAS 1982

```

RESULT 42

```

US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MUL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label=X
; OTHER INFORMATION: /note="X" = M and N, or N"
US-08-290-919-4
Query Match 3.3%; Score 282; DB 1; Length 53;
Best Local Similarity 94.2%; Pred. No. 7.1e-09;
Matches 49; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

RESULT 43

```

US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992

```



```

Db 1300 ETIELTAPKCKEMQNSLNDKNIYSIEIVVPKLLSDMLCQTEHLKTSKSTFLK----- 1355
Qy 1091 LKYYNGESSPLKLTSEESIQTEDNYSLEN-----FKVLSKLEGKLDNMLEKKLSYL 1146
Db 1356 -VKVHENVEKETAKSPATCYTQSPYSYIENSALAFYTSCSRSTYSQSLSLEAKK--WL 1412
Qy 1147 SSGHLHLIAELKEVINKNKYNTGNSPSENNTDNNMLESKKFLPBDTDAVATVSESGSDT 1206
Db 1413 RRG-----IEDGPERINTADYVGNLYENNSN-----STIAEDKNH 1450
Qy 1207 LEOQPPKPKPASTHVGASNTITTSQNVDEVDVILPIFGSEEDYDGLQGVVTEAVT 1266
Db 1451 LSEKQ-----DTYLSMSSMSNSYH-----SDEYVNSG----- 1480
Qy 1267 PSVIDNLIISKINEVEVLYLKLPLAGYRSLKQOLENNVTFVNVKDIINSRPNKRENF 1326
Db 1481 -----YLS-----KNKLDGSGIEPVKKNVEDQKNTSFSK----- 1508
Qy 1327 NVLESOLIPYKDLTSSNVYKDPYKFLNKKRDKFLSSVNYIKDSITDINFANDVLAGY 1386
Db 1509 -----VISN--VKDA-----NAYPQVNMEDICVEELVTS-- 1535
Qy 1387 KILSEKYSDDLSDIRKYYINDKOGENEKYLPLNNIET---LYKTVNDKIDLFYIHLEAKV 1443
Db 1536 ---SSPCKKNKNAIKLSTISNS-----NNEFVGPAPRIASGKI--VCVSHETIKK 1580
Qy 1444 LNTYTEKSANVEYKIKELNLTATLOKLDKFKNNNFVGIADISTDYNNHNLTKELST-- 1501
Db 1581 VMDITDTSFSKY--IKENNENKS--KICQTKIMAGCYEALDSEDLNHSLSLNDDECSH 1636
Qy 1502 GWFFENLAKTVSNLDGNIJGMLNISOHQCYKQCPQNSGCFRHLDEREPECKCLINTYQ 1561
Db 1637 HNVFADIQSEEL-----LQHNQNSGLEKYSKISP-----CDVSLSTIDICKCSIG--- 1682
Qy 1562 EGDKCVENPNPTC 1574
Db 1683 KLHKSVSAN--TC 1694

RESULT 46
US-08-639-501-2
Sequence 2, Application US/08639501
Patent No. 5837492
GENERAL INFORMATION:
APPLICANT: Tavtigian, Sean V.
APPLICANT: Kamb, Alexander
APPLICANT: Simard, Jacques
APPLICANT: Couch, Jacques
APPLICANT: Rommens, Johanna
APPLICANT: Weber, Barbara
TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., Suite 1001
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,501
FILING DATE: 29-APR-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/585,391
FILING DATE: 11-JAN-1996

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/576,559
; FILING DATE: 21-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/575,359
; FILING DATE: 20-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/573,779
; FILING DATE: 18-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Imnen, Jeffrey L.
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 24884-116802-04
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-962-4810
; TELEFAX: 202-962-8300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-639-501-2

Query Match 3.3%; Score 279.5; DB 2; Length 3418;
Best Local Similarity 21.0%; Pred. No. 1.6e-06;
Matches 330; Conservative 225; Mismatches 569; Indels 449; Gaps 81;

Qy 187 LRAKALNDVCAN-----DYCQIPENLK---IRANBLDVKLKLVFGY-----RKPL 228
Db 667 LRKCSNFTCSNNTVVISQDLRYEAKCNKEKQLQFTFPADSLSCLOEGOCENDPKSKYV 726
Qy 229 DNKDNV-----GANE--DYIKNNKTT---ENINELI--EESKTT-----ID 264
Db 727 SDIKEEVLAACHPVQHSVEYSDTDFQOSKSLYDHNASTLLPPTSKDVLNLMVTS 786
Qy 265 KKNNAKTEEEKKKLYQAOYDLSTYKOLEAHNLISVLKRIPTTLKNNENIK--ELL--D 320
Db 787 RKGESYKMSDKLK-----GNNYESDVELTKNI--PMKKNQDVCLANENKYNVELLPRE 837
Qy 321 KINEIKNPPANSNGTPTNLLDNKKKIEHEKEIKEIAK--TIKENIDSLFTDPLELEYVL 379
Db 838 KYRWVSPSRKVOFN-QNTNL--RVIQKNOEETTSISKITVAPDSEELFSD----- 885
Qy 380 REKNKIDISAVETKESPEPNPEYPTSYNDINNALNE---LNS---FGDLINP 432
Db 886 NENNEVFOY-----ANERNNALGNTKELHETDL--TCVNEPIPKNSTMVLVIGDTGK 936
Qy 433 FYTKEPSKN-IYTDNERKKFTNEIKETIR-KKKIESDKRSYEDR-SKSLNDITKEYE 489
Db 937 QATQVSIKKDLYVLAENK--NSVKQHIKMTLGQDLKDISLNDIKIPEKKNNDYNNKWA 994
Qy 490 KILNELYDSKPNNNIDLTPEKMMGRYSYKEKLTHTHTFSPSYENSKNLEKLTALAY 549
Db 995 GLGPISNHSFGGSF-----RTASNKEIKLSEHNIRKSKMFKFQIEEDYPTSLAC 1044
Qy 550 MEDYSARNIVKEKELKYK--NLISKIENELETIT-----ENIKKDEOLFPEK--TKDN 602
Db 1045 VE--IVNTIALDQKKLSFPQSTINTVSAHQSQVSVSDCKNSHTPOMLFSDKDFNSNH 1102
Qy 603 -RPDEK--ILEVSDIVKVOVQVYLMNKKIDELKQTOLIKNELKHNHIVPNSY----- 653
Db 1103 LTPSQKAEITELSTLEESGSQF-----EFTQRPKPSYILQ-----KSTFEVPEKQMTILKT 1154
Qy 654 -KOENKQEPYLYLVKK---ETDKLVKPKYE-----SLNEEKKNIKTGQSNSESP 703
Db 1155 TSEBCRDADLHVIMNAPSIGOVDSKQFEGTVEIKRKFAGLLKNDCKNSASGVLTDENEV 1214
Qy 704 STEGETIGQAT-----TKPGQOA---GSALDEDSVQAQOEDQKQAPVPVPVPEKKAQV 755
Db 1215 GFRGFYSAGHTKLNSTELQKAVKLFSDIENISETSAEVH-----PLSSSSKHDSY 1269
Qy 756 PTPPAPVNNKTENVSKLDYLEKLYEFLNFTSYCHKIYLVSHSTMNKILKQYK--ITKEBE 814

```



```

Db 1270 VSMFKIENHNDKVTSEKNNKCOL-----LIQNNIEMTGTGFVEEITENYKNKNTNED 1321
QY 815 SKLSSC-----DPLDLFNIONIPVMS-MF-----DSL 844
Db 1322 NKTTAARNHNLFPDSSSKNDYVCIHKDEDDLFTDHNICLKISGQFMKEGNQIK 1381
QY 845 NSISQL-FMEIYERKVCNLYKLKDNKIKNLEAKRVSTVKTSSSMQPLSLPOD 903
Db 1382 EDLSDLTFLEVARAKQACH-----GNTSNKEQL-TATKTEQNIKFETS----- 1424
QY 904 KPEVSANDDTSHSTNLSKLFEENILSLKNNNIYELLGOKSSEMFYERIKLSDTF- 962
Db 1425 -----DTEFQT-----ASGNISVAKESFNK--TVNFPDQPELHNF 1461
QY 963 -YNEFTNFKSKADDIN-SLNDESKRRKLEEDI-----NKLKTLQ 1002
Db 1462 LNSELHSDIRKNNMDLSTETDVKHKLKESVPVGTGMQVTFQOQPERDEKIKPTL 1521
QY 1003 LSEFLYNNKYLKL-----ERLFDRK-----KTVGRYKMQIKKLTLK 1039
Db 1522 LGFHTASGRKVKLAKESLDKVKNLFDKEGQTSFTHSQMAKTL-KYREACKDELEAC 1580
QY 1040 EOLE-----SKLSLNPKHYLQNFVYFENKKEAETA-EENTLENTKILKHYKG 1090
Db 1581 ETEITAPACKEMQNSLNDKMLVSIETVPPKLSLNDLCROTENLTKSIFLK----- 1636
QY 1091 LVRYNGESSPLTLSEESIQTEDNYASLEN-----FKVLSKLEGLKDNINLEKKLSYL 1146
Db 1637 -VAVHEVVEKETAQSPATCTCTNOSPVIENSALAFYTSQSRKTSVSQTSLEKK--WL 1693
QY 1147 SSGLHLIAELKEVYKNNYGTGSPSENMTDVNNALESYKFLPEGTVAIVVSESSDT 1206
Db 1694 REG---IFDQPERIMADYGVNLYENNEN-----STIENDKNH 1731
QY 1207 LEOSQPKRPASTVGAESNTITTSQNDVEDVDYIIPIGESREDDYDLGOVYTGAAVT 1266
Db 1732 LSRKQ-----DYLSNSSMSNSTYH-----SDEVNDSG----- 1761
QY 1267 PSYIDNLSKIENEYEVLYLPLAGVYRSUKOLENNVMTFNVKDIILSRKRENF 1326
Db 1762 -----YLS-----KNLDSGIEVLKVNEDQKNTSPSK----- 1789
QY 1327 NVLESLLPYKDLTSSNYVVKDPYKFLNKKRDKFELSSVYIKDSIDTDINPANDVLYG 1386
Db 1790 -----VISN--VKDA-----NAYPQTVNEDICVEELVTS-- 1816
QY 1387 KIISEKYKSDLSIKKYINKOGENEKYLPFLNNIET---LYKTVNKKIDLFVTHLEAV 1443
Db 1817 ---SSPCKNNMAIKLISLTS-----NNFEVGPAPFRASGKI--VCVSHETIKK 1861
QY 1444 LNTYEKSNVEVRIKELINYKTLTQDLADFKKNNNFVGIADLSTDYNNHMLTFLST-- 1501
Db 1862 VKRIFLDFSPKY--IKENENKS---KICQFKIAGCVAEALDDSDILHNSLNDCEGTHS 1917
QY 1502 GAVFENLAKTVLSNLDGNLQGLMNLISQHCYVKKQCPQNSGCFRHLDRECECKLNYKQ 1561
Db 1918 HKFVFAIQSEEI-----LQHNQNSGLEKVKISIP---CDVSLSTSDICKCSIG--- 1963
QY 1562 EGDGCVENPPTC 1574
Db 1964 KLHKSVSSAN-TC 1975

```

RESULT 47
US-09-044-946-2
Sequence 2, Application US/09044946
Patent No. 6033857

GENERAL INFORMATION:
APPLICANT: Tavligian, Sean V.
APPLICANT: Kamb, Alexander
APPLICANT: Simard, Jacques
APPLICANT: Couch, Fergus

```

APPLICANT: Rommens, Johanna
APPLICANT: Weber, Barbara
TITLE OF INVENTION: Chromosome 13-Linked Breast Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti
STREET: 1201 New York Avenue N.W., Suite 1001
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 22204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/044,946
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/639,501
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/576,559
FILING DATE: 21-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/575,359
FILING DATE: 20-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/573,779
FILING DATE: 18-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-116802-04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-044-946-2
Query Match 3.38; Score 279.5; DB 3; Length 3418;
Best Local Similarity 21.0%; Pred. No. 1.6e-06;
Matches 330; Conservative 225; Mismatches 569; Indels 449; Gaps 81;

```

```

QY 187 LIRAKLNDVCAN-----DYCOIPFNLR-----IRANELDVKLKLYEGV-----RKPL 228
Db 667 LRKSRNETCSNNTVISODLDYKAKCNKEKLOLFTTPPADSLSCLOEGCENDPKSKV 726
QY 229 DNIKDNV-----GKME--DYIKNNKTTI---ENINELI--DESKRT-----ID 264
Db 727 SDIKEVLAACHPVQSKSEYSDTPQSQSLKYDHEMNSTLILPTPSKDVLSNLVMS 786
QY 265 KKNKATKEEEKKLYAQYDLSTYNNKQLEBAHNLISLEKRIPTLKNNENK--ELL--D 320
Db 787 RGKESYKMSDKL-----GNYESDVELTKNI--PMERNODCALNENYKNVELLPPE 837
QY 321 KINETKNPPANGTNPNTLLDKNNKLEEHKEKEIKIAK-TIKPNIDSLTPDPLELYYL 379
Db 838 KYMRVAPSRKVOFN-ONTNL--RVLOKNOEETTSIKTIVNDSLELSD----- 885
QY 380 REKNKNIDISAKYETKESTPNEXYNGVYPLSYNDINNALNE--LNS-----FGDLINP 432
Db 886 NENNFVQV-----ANERNNLALGNTKELHETDL-TCVNEIIFKNSIMVLYGDTGDK 936
QY 433 FDTYKPSKN-IYDNERKKFINKIKIE-KKIESDKSYEDR-SKSLNDITKEYE 489

```


;; FILING DATE: 16-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/329,154
;; FILING DATE: 25-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/161,907
;; FILING DATE: 02-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/985,321
;; FILING DATE: 04-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/429,791
;; FILING DATE: 31-OCT-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: INGOLIA, DIANE E.
;; REGISTRATION NUMBER: 40,027
;; REFERENCE/DOCKET NUMBER: OPHD-01763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 705-8410
;; TELEFAX: (415) 397-8338
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2366 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-480-604A-10

Query Match 3.3%; Score 278.5; DB 1; Length 2366;
Best Local Similarity 18.5%; Pred. No. 1.2e-06;
Matches 347; Conservative 292; Mismatches 595; Indels 643; Gaps 92;

QY 121 DAKSADL-----KRRVRYLL-TIKELKYPQLPDLTNHMLCLCD-NIHGEYVL----- 167
DB 51 DINSLTDIYDTYKSGKNKALKKEKYLVELEKNNNLTPVKNLH-FWIGGQIND 109
QY 168 -----IDGEEINELLYKLNFEYD-----LLRAKLVDCANDYCOIPEFLKIRANELD 215
DB 110 TAININQOMKDVNS-DYVNVWFYSDNAPLINTLKKTVESALINDLLE-SFRENLDPRD 167
QY 216 VLKLIYEVYKRPDLNIRKNGKMEYTIKKN-----KTENINE 254
DB 168 YNK-----FFKRMELIYKQKNFIYKAKORENELIIDIIVKTYLSNEYKEIDELMT 223
QY 255 LIEESKTIIDKN-NATKEEEKKKLYOAOYDLSIYNKOLEAHNLISYLE-KRIDTIKKN 312
DB 224 YIEESLAKITONSGNDVNFEEFKNGES---FNLYEQELVERWNLAAASDILIRISALKEI 280
QY 313 ENI---KELL-----DKINEIKNPPANSANTPTLIDKNKKIEEH-----EK 352
DB 281 GGMVLDVMDLPGIOPDLFESTIEKPSVYVDPEWMTKLEIMKYKREIYEYTSHEHDMDE 340
QY 333 EIKE-----IAKTKFNIDSLFD-----PLELEYLRKKKNIDISAKVETKESTEPNE 402
DB 341 EVQSSFESVLASKDSKSEIFSLGDMEASPLEVK--IAFNSKGIINOGLISYKDSYCSNL 398
QY 403 YPNVG--TYPLSYNDINNALNELNSFGDLINPEDYTKEPSKIITYDNRKKINIKIE-- 458
DB 399 IVKQIENRKLTLNNSLNPALISEDNDFTTNTFIDISIAEAN--ADNR--FMMLGLGYL 454
QY 459 -----KIRIEKKIISDKKSYEDRSKSLNDITKEYEKLNEIYDSKFNNDITLNF- 509
DB 455 RVGFEPDVKTTINLSGPAAVAAVQDLD-----MFKEGSMNHLEADLRNEISIKTINS 509
QY 510 ---EKAMKRRSYKYEKLTHTHTFPASYENSKHNEKLTKALKYME-----DYSLRN 557
DB 510 QSTDEMASLWSPFDARAK-----AQPEYKRN-----YFEGSLGEDNDNLDFS-QN 554
QY 558 IVEKELIYYNLSKINIEIET---LVENIKKD-----EQLFEKKI 597
DB 555 IYVQKE--YLLEKISSLSRSEBGTIHYIVQOGDKISYEACNLPAKTPYDSVLFQKNI 612
QY 598 TKDE-----NRPDEKILE-----VSDIVKVOVQKV-----LIM 625

DB 613 EDSSEIAYYNNPDGGEIQEIDKYYKIPSIISDRKIKITFIQHGKDERNTDIFAGFDVDSL 672
QY 626 NKID---ELKKTQILIKVNLKHNHVPNSYQOE-NKQEPYLIYLUKEDIKLQKPMKV 681
DB 673 TEIEAIDLAKEDISPKSIEI--NLGCMFYSISINVEETYPGKLTKKDKIISELMPSI 730
QY 682 ---ESLI---NEEKKNITKYGOSDNPSEPEGETTGATKPPQQAQSALEGSVQAQAE 736
DB 731 SODSIIVSANQYEVNRINSGRRELLDHSGEW----- 761
QY 737 OKQAQPPVPVPEAKAQPFPAPAPYNNKTEENVSKLDYLEKLYEPLNTSYCHKYLIVSH 796
DB 762 -----INKEESIIK-DISKREYISPNP----- 782
QY 797 STMNEKILKQYKITEEESKLSLSCDPLDLFNIONNIPWYMSFSLNLSQLMEIYE 856
DB 783 -----KENKITVKS-----NLPELSTLQEIIRNNSSSDIELIE 817
QY 857 KEWV---CNLYKLDNDKIKNLPEAKKVT-----SVKTLSSSSMOPLSLTPQDK 904
DB 818 KVMTECEINIVSNIDQYVERIEBAKVLSDSINYIKDEKLIESIISDALCDLKQOE 877
QY 905 PE-----VSAND--DTSHTNLNNSLSKLEPNIISLGNKNIYQELIGOKSSENFYEKILKD 958
DB 878 LEDSHPISEPIDSEIDEGFSIRFINKEGTESJFVETEKITFSEY-----ANHTEBISKI 932
QY 959 SDFYNESFTNFKV-----SKADDINSLNDESKRRKLEEDINKLKTLO-----LSFDLYN 1009
DB 933 KCTIFEDTVNGKLVRKVNLDJTHEVNTLNAFPQSLIE-YNSKSELSNLSVAMKVOYVA 991
QY 1010 K-YKLEKLERLFPKRRKTVG-----KKMQIKKLT-- 1036
DB 992 QFSTGSLNTITPAKAVELVSTALDETDIDLPFLSGLPIATIIDGVSGLAKIKELSET 1051
QY 1037 ---LKEQLESKLN-----SLNPKH 1054
DB 1052 SDPLRQETEAIRIGIMAVNLTTATTAIITSLSGIASGFIILVPLAGISAGIPSLVNNEL 1111
QY 1055 VLOQNSV---FNNKKKEKIEIETEN--TLEMTKILKH-----YKGLVK 1093
DB 1112 VLDRKATKVDYF---KHVSLVETEGVFPLDDKIMPDODLVISEIDFNNSIYLGKE 1168
QY 1094 YNNGESSPLKLTSEE-----SIQTEBNVASELNFVLSLECKLKNL---NLEKK 1141
DB 1169 IYRMGGSGHTYDIDHFFSAPSTIYREPHLSTIYQVEVQKEELDLSKDLVLPAPNR 1228
QY 1142 KLSY---LSSGLHHLI---AELKEVIKKNKYGN-----SPSENNTDV 1178
DB 1229 VAWETGWTPLGRSLLENQGTKLDDRIRD-NYEGEYWRFAFIADALITTLKPREDTNI 1287
QY 1179 NNALESYKFLDEPQDVATVWSESGDPLESOPKKPASTHGAASNTITTSQ----- 1231
DB 1288 RINDLSNTR---SFIYPII---TTEYIRKLSYFSYSGGTVALSLSQYMMGAINI 1336
QY 1232 -----NVDEEDVYIYPIGSESEEDVDDGGVVTGAAVPPSYVDNLKINENY 1281
DB 1337 ELSQSDVMIIDVDNVRDVTI---ESDKIKKGDLEGLSTLSIEENKI--TLNSH 1387
QY 1282 EVLVLEKPLAGYRSLSKOLENNVMTFNVNVKDIILSRFNKRNFNVLSESDLI--PYRDL 1339
DB 1388 ELNFGGEVNG-----SNGFVSLTFSILEGINA-----IIEVDLSYSYKYL 1428
QY 1340 TS-----SNYYKD-----PYFLNKKKDKFLSYNTYKDSITDIDI 1376
DB 1429 ISGELKIILMNSNHIOQKIDYIGFNSLOKNIPYFVSQEGKEN-----GFINSTKGBL 1483
QY 1377 NFANDVLGYKLLSEKYSDLSI---KRYINDQGENEKEKLPPLNNIETLYKTVVNDKI 1432
DB 1484 -----FVSELPDVVLISKVYDDSK---PSFGYYSNNLKDVKVYTKQNV 1524
QY 1433 DLFVILHEAKVNLTYEKSNEVVKIKELNLYLKTIDOKLADFKNKNFV---GIADL----- 1485


```

OY 1037 ---LKEQLESKLN-----SLNNPKH 1054
||:||||:|
Db 1052 SPPILROELEAKIGIMAVNLTTATTAITSSIGIASGFIILVPLAGISAGIPSLVNNBL 1111.
OY 1055 VIQNFV---FENKKEAEIAETEN--TLENTKILKH-----XKGLVK 1093
||:||||:|
Db 1112 VLROKATKVVDF---KHVSLVETEGVFLLDDKIMMPODDLVISEIDFNNSIYLGKCE 1168
OY 1094 YNGESSPLKLTSE---SIOTEDNVASLENFKVLSLECKLKNL-----NLEKK 1141
||:||||:|
Db 1169 IPRMEGGSGHYTDDIDHFEFAPSITTYRPHLSIYDVEVQKEELDLSKDLVLPAPNR 1228
OY 1142 KLSY---LSSGLHLI---AELEKVIKKNKNTGN-----SPSENTDV 1178
||:||||:|
Db 1229 VFAMETGWTPLGRSLJENGTCKLLDRD--NYGEFTYWRFAFIADALITTLKPREDTNI 1287
OY 1179 NNALESYKFLPEGTAVATVSESGSDTLEOSOPKRPASTHVGAESNTITTSO----- 1231
||:||||:|
Db 1288 RINLSMNR---SFIVPII---TTEYIREKLSYFSGSGTYALSLQYNNGINI 1336
OY 1232 ---NYDDEVDIYIPIFGESEEDYDDLQVYVYTGAVIPPSVDNLKIEHY 1281
||:||||:|
Db 1337 ELSSEDMVIDVNVVADVTI---ESDKIKKGDLLEGILSTLSEENKI--ILNSH 1387
OY 1282 EYLYLKPLAGYRSLKOLENNVMTFNVVKDILNSRFNFKRNFKVLESOLI--PYKDL 1339
||:||||:|
Db 1386 EINFSGEVNG-----SNGFVSLTSPSILEGINA-----IIEVDLSYKLL 1428
OY 1340 TS-----SNVYKDD-----PYKFNKREKDKFLSYNYIKDSIDTDI 1376
||:||||:|
Db 1429 ISGELKILMLNSNHIOQKIDYIGFNSLOKNIPYSFVDEGKEN-----GFIENGSRBEL 1483
OY 1377 NFANVNLGYKRLSKYSDDSI---KKYINDOGENEKYLPELNIETLYKYVNDKI 1432
||:||||:|
Db 1484 ---FVSELPDVLVLSKRYVMDSK---PSFGYYSNNLKDKVKTTKDMV 1524
OY 1433 DLFIHLBAKVLNYYEKSNEVEKIKELNLYKTIODKLADFPKNNNFV---GIADL--- 1485
||:||||:|
Db 1525 N-----ILTYGLKDDIKISIS-----LTLODE--KTILNSVHLDSEGAELKWM 1569
OY 1486 ---STDYNNHNLTKFLST---GMVEFENLAKTVLSNLDOGN--LOGLMNLISQHCYKQOP 1538
||:||||:|
Db 1570 NRKGTNTSDLSMFSLESMINKISIFVNFLOSNIKFIIDANFIISGTSIGQEFPI---CD 1626
OY 1539 QNSGC-----FRHL-----DERECKCLNRYKQEG---DK 1565
||:||||:|
Db 1637 ENDNIQPYFIKFNLTETNYTLYVGNRQNMIVEPNYDLDDSGDISSTVINFSOKLYLGIDS 1686
OY 1566 CVE---NPNPTCNENN 1578
||:||||:|
Db 1687 CVNKVVISPNITDEIN 1703

```

```

ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,136
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/480,604
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 397-8338
TELEFAX: (415) 705-8410
INFORMATION FOR SEO ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
type: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-10

Query Match 3.3%; Score 278.5; DB 4; Length 2366;
Best Local Similarity 18.5%; Pred. No. 1.2e-06;
Matches 347; Conservative 292; Mismatches 595; Indels 643; Gaps 92;

OY 121 DAKSYADL-----KHRVRYLL--TIKELKYPOLFDLTNMLTLC--NIHGFYLL----- 167
||:||||:|
Db 51 DINSLDIYIDTYKKSGRKALKKRFREYLVTEVLELKNNNLTPVEKNLH--FWMIGQIND 109
OY 168 -----IDGYEELNELLYKINFED-----LLRAKINDVCANDYCOIPFNLIKIANELD 215
||:||||:|
Db 110 TAINIYNQKDVNS--DYNNVNFYDSNAFLINTLYKTKVESAINDTLE--SFRENLDPRD 167
OY 216 VLKTLVFGYRKPLDNIKQNVGKMEYIKKN-----KTIEINE 254
||:||||:|
Db 168 YNK-----FFRKMEITIIYKQKNFINYKAKQRENEBELIIDIVKYTILSNYSKEIDELWT 223
OY 255 LIEESKTIIDKNK--NATKEEKKRLYQAOYDLSIYNKOLEAHNLISYLE--KRIDTLKN 312
||:||||:|
Db 224 YIEESLNTKTONSGNDVRFEEFKNGES---FNLYEQELVERMNLAAASDILIRISALKEI 280
OY 313 ENI-----KELL-----DKINEIKNPPRANGNTPNLLDKNKKIEEH-----EK 352
||:||||:|
Db 281 GGMVLDVDMPLGIQPDLEESIEKPSVYVDEWEMTKLEAIMYKKEYIPEYTSHEHMDLME 340
OY 353 EIKE-----IAKTIKFINDSLFTD-----PLELEYLRKNKNINDISAKVETKESTEPNE 402
||:||||:|
Db 341 EVQSSFESYLASKDSKSELFSSLGMEASPLEYK--IAFNSGIINOGLISKDYCSNU 398
OY 403 YPNGV--TYPLSYNDINNALNELNLSFGDLINFDYTKPSKNIYTDNERKKFINIKE-- 458
||:||||:|

```



```

Db 399 IYKQIENRYKLLNNSLNPASEDNDFNTTNTFIDSIMAEAN--ADNGR--FMMELGKYL 454
QY 459 -----KRIEKKIESDKKSYEDRSKSLNDITKEYEKLNLEIYDSKRNNDLTFN- 509
Db 455 RUGFPDVKTTILNSGPEAAAYODLL-----MFEKGSNNIHLIADLRNFEISKTNIS 509
QY 510 ---EKMGRKRYSYKVEKLTJHNTFASYENSHNLEKTLKALYME-----DYSLRN 557
Db 510 OSTEQEWASLMSFDDARAK-----AOFEERYKRN-----YFEGSLGDDNDLDFS-QN 554
QY 558 IYVEKELKYKKNLSIKIENIET-----LVENIKKD-----EOGLPEKKI 597
Db 555 IYVDKE--YLEKRISSLARSSERKGYIHYIYOQGDKISYEACNLAKPTDYDVLQKNI 612
QY 598 TKDE-----NKPEKILE-----VSDIVKVOQKV-----LML 625
Db 613 EDEIENAYYVNPBGGEIOEIDKYKIPISIDRPKIKLFIHGKDEFTDIFAGEVDVSL 672
QY 626 NKTD---ELKKTOLLKKNVELKHNHVPNSYKOE--NKQEPYLYLVKKEIDKLKVPKPV 681
Db 673 TEIEAIDIAKEDISPKSIEI--NLGCMNFSYSINVEETYPGKLLKVKDKISELPSI 730
QY 682 --ESLI--NEEKKNKITEQSDNSEPTEGELTGQATKPGQAQSGALGDSYQAQOE 736
Db 731 SODSIIVASAOYEVRIINSEGRRELDHSGEW----- 761
QY 737 OKOQOPVPVPEAKAOYPTPPAPVNNKTEVNSKLDYLEKLYEFLNTSYCHKYLIVSH 796
Db 762 -----INKESIIR--DISKEYISNRP----- 782
QY 797 STYNEKILKOYKITEEESKSLSCDPLDLLFNIONNIPWYSMFDLSNLSOLEMEIYE 856
Db 783 -----KENKITVYKSK-----NLPELSTLQOEIRNNSNSDIDLEEE 817
QY 857 KENV-----CULYKLNKDKIKNLEEKAKVST-----SVTLSSSSQPLSLTPQDK 904
Db 818 KVLTECEIENIVSINIDQIYEEKREKKNLTSOSINYIKOEKFLIESISALCDLKQONE 877
QY 905 PE---VSAND--DTSHSTLNNSLKLEENILSGKNKNIYOLIGKSSSENEFEKILKD 958
Db 878 LEDSHFISFEDISETDGBFIRFNKTEGSIYETEKTIFFSEY-----ANHTEIISKI 932
QY 959 SDFPFYNSFTNFKV---SKADDINSUNDESKRKKLEEDINKLKTIO-----LSFPLYN 1009
Db 933 KGTIFPTVNGKLVKKNLDTTHEVNTLNAAFIOSLIE--YNSSKESLSNLSVAMKVOYVA 991
QY 1010 K-KYKLEERLFDKKKTG-----KXKMQIKKLT-- 1036
Db 992 QLFSTGLNTTIDAKVVELVSTALDETIDLLPILSEGLPIIATIIDGVSIGAAIKELSET 1051
QY 1037 ---LKEQLESKLN-----SLNPKH 1054
Db 1052 SDPLLKQEIKAIGIMAVNLTATATITSSLGASGFTLVLPLAGISGIFSLVNNEL 1111
QY 1055 VLONFSV---FFNKKKEAIEAETEN--TLENKILILKH-----YKGLV 1093
Db 1112 VLKDKATKVVDF---KHVSLVTEGEGVFTLLDDKIMMPQDVLVSEIDFNNNSIVLCKCE 1168
QY 1094 YVNGESSPLKLTSEE-----SIOEDNYVASLENKVLSKLEGGKDKUL---NLEKK 1141
Db 1169 IMWEGSGSGHTVTDIDHFFSAPSTIYREPLSTIYDYLEVQKEBLDSLKDLMLVLPNPNR 1228
QY 1142 KLSY---LSSGLNHLI---AELKEVINKNVYTGK-----SPSENNTDV 1178
Db 1229 VFAMEIGMTGLSLSDNGKLDLRID--NYEGEFYKRYFAFTADALITLTKPNEYDTNI 1287
QY 1179 NNALESYKKELPFGTDVATVVSSESDTLEQSPKRPASTHVGAESNTITTSQ----- 1231
Db 1288 RIMLDSWTR-----SFLVPII-----TTEYIREKLSYSFYSGGYALSLSQYMGGINI 1336
QY 1232 -----NVDDEVVDVLIIVIFGSEBEDYDLDGOVTVGEAVTPSVIDINILSKINEY 1281
Db 1337 ELSSEDMWIIDVNVADVTI-----ESDKIKKGLDLEGILSTLSIEENKI--ILNSH 1387

```

```

QY 1282 EVLKPLAGVYSRSLKKQLNNVNTFNVNVKDLNLSRFRNKRREKNVLESOLI--PYKDL 1339
Db 1388 EINFSGEVNG-----SNGFVSLFSLLEGINA-----ILEVDLSSKSVKLL 1428
QY 1340 TS-----SNVYKD-----PKFLNKRKDKFLSSYVITDSDTDTI 1376
Db 1429 ISGELKILMNSHNHIOKIDYIGFNSLOKNIPIYSPFSDSGKEN-----GFIGSTKEGL 1483
QY 1377 NFANDVLGYKKILSEKYSKSLDSI-----KKYINDKOGENEKYLPLPNNIETLYKTVNDKI 1432
Db 1484 -----FVSELPPVYLISKYIMDSK---PSFGYSNNLKDVKVITKDNV 1524
QY 1433 DLFVHLEAKVLYTYEKSANEVAKIKELNYLKTIOKLADEFKNNNFV---GIADL--- 1485
Db 1525 N-----ILNGVYLKDDIKISLS-----LTLQDE-KTIKILNSVHLDSEGAELIKFM 1569
QY 1486 --STDYHNNLLRFLST--GWFEEFLAKTVLSNLLDGN--LQGMNISOHCYVKQCP 1538
Db 1570 NRKGNMTFSDLSLFSLSMNIKSIIVNFLOSNIKFIIDANFITSIGTYSIQOFEPT--CD 1626
QY 1539 QNSGC-----FRHL-----DEREDECCLANYKQEG---DK 1565
Db 1627 ENDNIOPIFKFNTLENTLYLYGNRQNMIVEPNYDLDDSGDLSISTVINSQKYLIGIDS 1686
QY 1566 CVE---NPNPTCENN 1578
Db 1687 CVNKVVISPNITYDEIN 1703

RESULT 52
US-08-957-310-10
; Sequence 10, Application US/08957310
; Patent No. 6365158
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPMD-01121
; TELECOMMUNICATION INFORMATION:

```


TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-310-10

Query Match 3.38; Score 278.5; DB 4; Length 2366;
Best Local Similarity 18.5%; Pred. No. 1.2e-06;
Matches 347; Conservative 292; Mismatches 595; Indels 643; Gaps 92.

QY 121 DAKSADL-----KRVNRYLL-TIKELKPOLPDLTHMLTCD-NHGRKYL----- 167
DB 51 DINSTLDYIDYKSGSKKALKRKEKLYLEVLELKNNNLTPVKNKH-FWIGQIND 109
QY 168 -----IDGVEEINELLYKLNFEVD-----LLRAKLVDCANDYCOIPLNKKIRANEID 215
DB 110 TAININOMKDVNS-DYVNVVFDYDANFLINTLKTIVYESAINDTLE-SFRENLDPRFD 167
QY 216 VLKRLVGRKRPDLNIDKDVGMEDYIKRN-----KTIENINE 254
DB 168 YNK-----FFRRKMEIYDKQKNFIYKKAOREENPELIIDIIVKTYLSNEYSKEIDEINT 223
QY 255 LIEESKRTIDNK-NATKEEKKKRLYOAOYDLSIYNKOLEEHNLIISVLE-KRIDTLKN 312
DB 224 YIEESLKNITONGNDVNFEEFKGES--FNLYEDELVERMNLAAASDLIRISALKEI 280
QY 313 ENI-----DKINEIKNPPRANSNTPTLLDNKKKEEH-----EK 352
DB 281 GGMVLDVDMLPGLDPLESIEKPSVTVDFEMWTKLEAIKKRYIEGYSEHMDIDE 340
QY 353 EIKE-----IAKTKEINISLFTD---PLELEYLREKKNKNDISAKVETKESTEPNE 402
DB 341 EVOSFESFVSLAKSDKSEIFSLGDMEASPLEVK--IAFNSGIIINOGLISVYKDCYSNL 398
QY 403 YPNGV--TYPLSYNDINNALNELNFEGLINPEDTKEPSKIYTDNRRKFINIKE-- 458
DB 399 IVKOENRKILNNSLNALISEDNENTTTTFIDISIMAEAN-ADNCR-FMMLGLKYL 454
QY 459 -----KIRIEKKIISDKSYEDRSKSLINDITREYKELINETYDKFNNNIDLTNF- 509
DB 455 RVGFPPDVKTTINLSGPEAYAAAYODLL--MFKEGSMNHLEADLRNEISIKTNIS 509
QY 510 ---EKMGKRYSYKVEKLTHTNTFASYSNKHLEKTLKALYME-----DYSLRN 557
DB 510 QSTDEMASLWSPDARAK-----AOFEYKRN-----YFEGSLGEDDNIDLFS-QN 554
QY 558 IVEREKELYYKLNLSKINEIET---LVENIKRD-----EEOLEFEKI 597
DB 555 IYVDRK--YLEKISLSLARSSERGIYIYVLOGDKISYEAACNLFAKTPYDSVYLQKNI 612
QY 598 TKDE---NKPDEKILE-----VSDIYKVOYORV-----LLM 625
DB 613 EDSEIAYYNGDGEIQEIDKYKIPSIISDRPIKILTFIIGKDEFNDFIDFAGFVDLS 672
QY 626 NKID---ELKQTOLIKANVELKHNTHVNSYKQF-NKOEPIYLLVLAKEIDKALFKPMKV 681
DB 673 TEIEAIDLAKLEDISPKSIEI--MLGCMNFSYSINVELEYPGKLLLVKVKISIELMSI 730
QY 682 --ESLI---NEEKNIKIKESGDSNESPTEGEITGQATTKPQOQGSALGDSVQAQOE 736
DB 731 SODSIIYSANQYEVAINSEGRRELLDHSGE-- 761
QY 737 QKQADPPVVPVPEAKAQPPTPAPVNNKTENVSKLDYLEKLYEPLANTSYICHKYLIVSH 796
DB 762 -----INKESEIIR-DISSEKISFPN----- 782
QY 797 STMNEKILKQYKITEESKSLSCDPLDLFRINQINNIIVMTSMFSLNNSLSQLEMEIYE 856
DB 783 -----KENKTIYKSK-----NPELSTLLLOEIRNNSNSDIELEE 817

QY 857 KEMV-----CNLYKLKDNKIKMLLEAKKVS-----SVKTLSSSMQPLSLTPDOK 904
DB 818 KWLMECEINLVNSIDTOIVERIEEAKNVLSDSINYIKDEKFLIESIDALCDLKQONE 877
QY 905 PE---VSAND--DTSHTNLSNLSKLEFENILSLCKNNIYQELIGKSSSENFYEIKLD 958
DB 878 LEDSHFISFEDSETDEGFSIRFINKEGTGESIFVETERTIPSEY-----ANITFEISKI 932
QY 959 SDTFYNESFTNFKV---SKADINSLNDESKRKLLEEDINKKTKLO----LSFDLYN 1009
DB 933 KGTIDPTVNGKLVKRVNLDTHENVTLNAAFTIGLIE-YNSKSELSNLSVAMRVQYVA 991
QY 1010 K-YKAKLERLPDKKTYG-----KYKQKIKL-- 1036
DB 992 QLESTGLMTITDAKRVELVSTALDETIDLLPTLSEGLPIATIIDGVSLGAIKLESET 1051
QY 1037 ---LKEOLESKLN-----SLNPKH 1054
DB 1052 SDPLLRQLEAKIGIMAVNLTATTAITTSIGIASGFSILLVPLAGISAGIPLSVNNEL 1111
QY 1055 VLQNFV---FENKKKAELIETEN--TLENTKILKH-----YKGLVK 1093
DB 1112 VLRDKATRVVDYF---KHVSLVETEGVFTLLDDKIMMPQDDLVISEIDFNNSIYLKCE 1168
QY 1094 YNNGESSPLKTLSEE-----STQEDNYASLENFVLSLEBKLDONL---NLKK 1141
DB 1169 IWRMGSGGHVYTDIDHFFSAPSTTYRPHLSITDYLEVQKEEILDLSKDLVLPNAPNR 1228
QY 1142 KLSY---LSSGLHLL---AELKEVIKKNKYTG-----SPSENNTDV 1178
DB 1229 VFAMETGWTGRLSLLENQOTKLLDIRD-NYGEGERYWHYFARIALITTLKPREYDINI 1287
QY 1179 NNALESYKFLPECTDVATVVSSEGSPTLEOSQPKPASTHVGASNTITTSQ----- 1231
DB 1288 RLNDSMNR--SFIVPII--TTEYIREKLSYFSGSGGYALSLSQYMMGINI 1336
QY 1232 -----NVYDEVDVYIPIVGESEEDVDLGOVATGAVPVSVIDNLSIKENEY 1281
DB 1337 ELSESDWMLIDVNVYRDTI-----ESDKIKKGDLIEGLISTLSEENKI--ILNSH 1387
QY 1282 EYLYLKLPLAGYRSLSKOLENNVMTFNVVNDILNSRNNKRENFKNVLESDLI--PYRDL 1339
DB 1388 ELNFGGEVNG-----SNGFVSLTFPSILEGINA-----ILEVDLSKSKYL 1428
QY 1340 TS-----SNYYKD-----PYFLNKKERDKFLSSYNTIKDSIDIDI 1376
DB 1429 ISGELKILMNSNHIOOKIDYIGFNSLOKNIPIFSVDSSEKEN-----GFLNGSTKGL 1483
QY 1377 NFANDVLGYKILSEKYSDDLSI---KRYINDQGENEKYLPPLNNIETLYKTVNDKI 1432
DB 1484 -----FVSELPDVLISKYVYDDSK--PSRGYSNNLKDVKYVTKRNV 1524
QY 1433 DLFVILHLEAKVLYNTERKSNVEVKIKELNLYKTTIDOKLADFPKNNPV--GIADL--- 1485
DB 1525 N-----ILTYGLYKDDIKISLS-----LTLQDE-KTIKLVNHHDESVAIELKFM 1569
QY 1486 ---STDYNNNNLLTKFLST---GMVPENLAKTYLSMLDGN--LQGLMLNISQHCYKQKCP 1538
DB 1570 NRGKMTNTSDLSMFLSEMNIRKISIFVNLQSNIKFIIDANFIISGTSIGQEFI---CD 1626
QY 1539 QNSGC-----FRHL-----DERECCCLANYKQEG---DK 1565
DB 1627 ENDNIOPIFKENTLETWTLYLVGRQNMIVPEPNYDDDDSGDISSTVINFSGKLYXGIDS 1686
QY 1566 CVE---NPNPTCENN 1578
DB 1687 CVNKVVISPNITYTDEIN 1703

RESULT 53
US-09-336-447A-76
; Sequence 76, Application US/09336447A

Patent No. 6310190
GENERAL INFORMATION:
APPLICANT: HANSEN, ERIC J.
APPLICANT: AEBI, CHRISTOPH
APPLICANT: COPE, LESLIE D.
APPLICANT: MACIVER, ISOBEL
APPLICANT: FISKE, MICHAEL J.
APPLICANT: FREDENBURG, ROSS A.
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
FILE REFERENCE: AMCY:024
CURRENT APPLICATION NUMBER: US/09/336,447A
NUMBER OF SEQ ID NOS: 98
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 76
LENGTH: 3788
TYPE: PRF
ORGANISM: Moraxella catarrhalis
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (1036)..(3786)
OTHER INFORMATION: Xaa = any
US-09-336-447A-76

Query Match 3.3%; Score 278.5; DB 4; Length 3788;
Best Local Similarity 17.2%; Pred. No. 2.1e-06;
Matches 307; Conservative 290; Mismatches 689; Indels 495; Gaps 72;

QY 32 LKLEAVLTLGYLSPQKEM-----VLNBTSGAVTTPPGSGSVASGSGSVA 83
DB 1258 INLESNEBGLXXLSQSDINPRTTEINLEARNAXA-NYATPSTINSALPENGSEQ 1316
QY 84 SGG-----SVASGSGSVASGSGSRTNPSNDSDDAKSYA----- 126
DB 1317 IDNPREINLEARNASTDGLKXVSEQDINPRTTEINLEARNAXANATPSTINSI 1376
QY 127 -DKHVRNLTLEKELKLPOL-FDLTN-----MLTCDNIGCFKLIDYE- 172
DB 1377 TDLGTIVDFEXXSEQIDNPREINLEARNAXANATPSTINSITDGLTIVDSEQI 1436
QY 173 -----EINELLYKLFYFDLRAKLN-----DYCANDYCOIPLMKIRANLVDLKLKLVFG 223
DB 1437 DNPREIN-LEARNANVADLTKVNALDKVASDXTSEQIDNPREINLEARNAXAN 1495
QY 224 YRKLPLNDKDNVGMEDYIKKRNKTENINLEIESK---KITDKKNATKEE---EKK 276
DB 1496 YATPSTINSALAEQ-----LNGKTLTPVSEQIDNPREINLEARNAXAKHDASTEGK 1551
QY 277 KLYQADYSTYKOLE-----EAHN--LISYLEKRIIDLKKNENIKELDKINETKNP 329
DB 1552 KMSQID-----NPREINLEARNAALESNVEEGLDLSGSEQIDNPREINLEARN 1607
QY 330 PANGMTPTLDDKKNKIEHEKEIKEIKTIKFNIDSLPTDPLEEYLYREKKNKNDIS 389
DB 1608 NANO-----NLIKRTANKSE-----QID-----NPREIN 1633
QY 390 AKVETESTEPNEYPGVTVPLSYNDINNALNLESGDLINPDTYKEPSKIYTDNR 449
DB 1634 LINEARNALDKNE-----YSKSEQIDNPREINLEARNNA-SITDGLTKSEQIDNPR 1686
QY 450 KR--FINEIK--EKIKIEK-KKIESDKSY---EDRSLSLDITKEYEKLINETYDSK 499
DB 1687 TEINLEARNANQNTLESEQIDNPREINLEARNALAE--QOLETTLKSEQIDN 1744
QY 500 FNNIDLTNEKMMGRYSKYV-----KLHNTFASVENSKNHLEKTLKLYMEDY 553
DB 1745 PRTEINLEARNANSDSQIDNPREINLEARNANADADASETLTKSEQIDNPR 1804
QY 554 SLRNIYVERELKYYKNLISIEIETIVENIKDESOLEFEKIKTDENKPDKILYEVSD 613
DB 1805 TEINLEARNANAFATANDKSE---QIDNPREINLEARNANASSNTQNIKASEQ 1861
QY 614 IYKVOYOKVLLMKNIDELKKTQTLKLVNELKNIHVNSYKQENKQEPYLLIVLKREIDK 673

DB 1862 IDNPRE-----INLEARNARLLDQSEQIDNPREINLEARNANATADATKNGXS 1917
QY 674 LKVFMPKVE-SLINEERKNK-TEGQSDNSE-----PSTGEITGOATTKGQOAGSALG 727
DB 1918 EQIDNPREINLEARNANAKAXANDRSEQIDNPREINLEARNANANATPSTITI 1977
QY 728 DSVQAO-AOEQOQAPVPVPEAKAQPPTAPVYN--NKTENVSKLDYLEKLYEPLNT 784
DB 1978 NSNQADIAQOQTQIDLAANELOSE-QIDNPREINLEARNANQAD----- 2025
QY 785 SYCHKIYLVSHSTMEKILKYKIKTEESKLSQDPLDLNIONNIPVMSMPSLSLN 844
DB 2026 -----LANNT----- 2030
QY 845 NSLSQFMEIYEKENVCLYKLNNDKIK-NLLEAKKYSTSVKTLSSSSMQPLSLTPD 903
DB 2031 NNITELAQOQDQSEQI-----DNPREINLEARNANVNEQOT--EALDALNSEQID 2080
QY 904 KPEVSAN-----DTSHTNLNLSLKELENILSLGKNKNIYOELLIGKSSENFYE 953
DB 2081 NPRTTEINLEARNALIGDAIVNSQDS-----EQIDNPREINLEARNAKALESNVE 2136
QY 954 KILKSDTFYNESFTNFWVSKADDINSLNDESKRKLEEDINKLTKLOLSEPLYNKYKL 1013
DB 2137 EGLIDL-----SGRSQIDNPREINLIN-EARNAALESNVE--EGLLELS----- 2179
QY 1014 KLERLFQKKTGKVMQIKKTLLEQLESKLSLN-----NPKHVLQFVFFN--KK 1066
DB 2180 -----GRTIOQSEQI-----DNPREINLEARNANQAHANNIKNIYELAQ 2224
QY 1067 KEAEIAETENTLENTILKLVKGLVKKYNGESSPLKTISESIQTEDNYSLENFKVLS 1126
DB 2225 IDQSEQIDNPREINLEARNAXANY-----ATPSTINQA-DIAQOQTQIDLAAYN 2279
QY 1127 KLEGLKLDN-----LNEKKLSTLSSGLHLINLEKEVYKKNKYTGNSPEENNTDVNNAL 1182
DB 2280 ELQSEQIDNPREINLEARNATHDYNERQTEASEQIDNP-----RTEINLIN 2329
QY 1183 ESYKFKLESTGDAVAVVSEGSDTLEQSQPKKASTHVAESNTITTSQNVNDEVDVII 1242
DB 2330 EARN-----AKASSENTONIAKSEQIDNRT-----EINLEARNAA----- 2366
QY 1243 VPIFGESEDDYDLGOVYVTEAVTPSYVIDNLSKINENEYVLYLKLPLAGYRSKLQLEN 1302
DB 2367 -----MILQDT--AIVSN--SODNKIQOLKEFK-----SEQIDN 2395
QY 1303 NWTFNPNVNDILNSRFNKNREKNVLESOLIPYKDLTSSNYVVKPYKFLNKEKDKFL 1362
DB 2396 P-----RTEINLEARNAGOTIIPLDQXAPSEQIDNP-----R 2431
QY 1363 SSYNYIKDSIDTINFANDVLGYKILSEKY---KSD-LDSIKRYINDKGENEKYLPFL 1418
DB 2432 TEINLEARNANANATPSTITNSLHEQOLKGSQIDNPREIN-----LINEA 2483
QY 1419 NNIEYLYKYNDKIDLFVHLEAKVLNITYEKSNEVEKIKELNYLTKTIDOKLADFK-- 1475
DB 2484 RNAXANATPSTINIIF-----FNXSEQIDNPREINLEARNAXANVATPSTI 2533
QY 1476 --NNNFVGIADLSTDYNNHNLTKPLSTGMVFENLAKTVLSNLDGLQGLMISQHCY 1533
DB 2534 TNNNNNNIYELAQOQDQSHSDIKTLSEQI--DNPREINLEARNAGTGCAGTCA 2590
QY 1534 KQCPQNSGCFRLDE-----REB--CKC-----LANY 1559
DB 2591 GATCAGTGACSEQIDNDNALINERDBLEBCCACCAACGACGACSEQIDNDNALIN 2650
QY 1560 KQF-----GDKCVENPNTCENN-----GGC 1581
DB 2651 ARDBLEAGCGGTCCGTGATGACGSEQIDNDNALINERDBLCGTATCAAGCAGCG 2710
QY 1582 DADAKCTEEDSGSNGKKIT-----CECT 1604

TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEFAX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label=X
OTHER INFORMATION: /note="X = M and N, or N"
US-08-290-919-1

Query Match 3.3%; Score 274; DB 1; Length 48;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1527 ISOHCVKRCQPNCGCFHLDRECKCLLANKOGDKCVENPPT 1573
DB 2 ISOHCVKRCQPNCGCFHLDRECKCLLANKOGDKCVENPPT 48

RESULT 56
US-08-685-871-2
Sequence 2, Application US/08685871
Patent No. 6013499
GENERAL INFORMATION:
APPLICANT: NARUMIYA, Shuh
APPLICANT: IMAMATSU, Akihito
TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,871
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184102
FILING DATE: 25-JUN-1996
PRIOR APPLICATION DATA: JP 7-262553
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-685-871-2

Query Match 3.3%; Score 274; DB 3; Length 1354;
Best Local Similarity 19.6%; Pred. No. 1e-06;
Matches 239; Conservative 195; Mismatches 389; Indels 394; Gaps 56;

QY 412 SYNDINNALNELNSFGDLINPPFYTKPSKNT---YTDNE---RKKFINEIEKIEK 464
DB 286 TYSKINNHKNSL-TFPD---DNDISKE-AKNLICAPLTREVALGRNGVEIKRHL--- 336
QY 465 KRTESDKSYEDRSKSLNDITREKELNLIYSKNNNDLNF---EKKMGKRSYKY 521
DB 337 -FFKNDQMAWEITRTDTAARVPD-----LSSDIDISNFDLEDKGEETFP 383
QY 522 EKLTHNT---FASYSNSKH-----NLEKTLALYMEDYSLRNIVKELKYRN 569
DB 384 PKAFVGNQLPEFVGFTYSNRRYLSANPDNRRSNADKSLOE-SLOKTIYKLEPDLNE 442
QY 570 LISKINEIETLVENIKKDEOLFEEKTKR--DENKPDKEILEVSQIVVQVYLLMK 627
DB 443 MOKLDEMEQKCRSTNKLKD-----KIMKEIDECNQRNLE-STVSQIEKEMLLQHR 494
QY 628 IDELKTQULIKVLELHNHIVPNYSKOENKOPRYLLYLKKEIDKLKVFMPKVESLINE 687
DB 495 I-----NEYORKAEQD-----NE 507
QY 688 EKKNIKTEGQSDNSEPSTEGEITGOATTPRGQAGSALGDSVQAQAOQKQAPVPVP 747
DB 508 KRRNV-----EREVS-----TLKDQLEDLKKV----- 529
QY 748 VPEAKQVPTPPAPVNNKTEVNSKLDYLEKYLEFNTSYICHYIIVYSTNMEKILKQY 807
DB 530 --SQNSQL-----ANEKLSQLOK--QLEBANDLTFESDTAVRLKSHTEKSKSLSQLE 579
QY 808 KITKEESKLSGCDPLDLFNIONNIPVMSMEDSLNNSQLQFMEIYEKEMVCNLYKL- 866
DB 580 SLNRELQER-----NRI-----LENSQSYDXYOQALILEARRD 616
QY 867 --KNDKIKNLEAKKAVSTSVTSSSQMLSTLPQDKPEVSANDTSHSNLNSLK 924
DB 617 RGHDSSEMIGLQARITSLOEVYHLKH-----NLKVEBERKQADMLNSEKKNLE 670
QY 925 LFNIIISLGKKNIIYOELGOKSSENFYEKLKDSOTFYNESPTNPKSKA--DDINSLN 982
DB 671 IDLNT---KLSLQORL--QEVNEH---KVTKARLTDKHQSLF-AKSYAMCEMKKLL 721
QY 983 DESK-RKLEEDINKLKLTL-QLSPDLYNKYKLLEPDKTKVYKQMKIKLTLKE 1040
DB 722 EEREAREKAKENRVYQIEKQSMLDVDL-KSQQKLEHLGNKE--RMEDEVNLTLOLE 777
QY 1041 QLESKLSLNNPKHVLYQNSVFENKKEKAEIATEVNTLENTKTLK-LHYGLVYKYNES 1099
DB 778 QESNKRLLLONELKT-QAEFADNLKLEKQKQEIITLLEAKRLLEFELAQLTKQYRNGE 836
QY 1100 SPLKTTSEESIQTEEDVYASLENFKVLSKLEGRKQDNLNLEKRLSYSSGLHLLIALEKE 1159
DB 837 GQNR-----ELQDQLEAQ-----YFSTLYKQVAKLEKE 865
QY 1160 VIKNNKYTGNSPENNTDVNNALLESYKFLPECTDVATVVSSES---GSDTLQO-----S 1210
DB 866 EIEEKN-----RENILKKIOELQNEKETLATQDLAETKAESEQALRGLEEQYELQ 918
QY 1211 QPKKPASTHVGAESNTITTSQNVDEVVDYIIVPIGSESEDDDDLGQVYTGAVVPSYI 1270
DB 919 ESKKAASRN-----KQELTIDK----- 934
QY 1271 DNLSKIENEYEVLYLPLAGVYRSKLQLENNVMTFENVVKDILNSRFKRNFKNVLE 1330
DB 935 DRTVSLKE-----ANSMFL-----KDI-----ELIKREN----- 959
QY 1331 SDLIPYKDLTSSNYYVYKDPKFLNKKERDKFLSSYNYIKDSITD-----INFANDYL 1383
DB 960 -----EELTEKKMKKAEEEKKL--EKEEISNLKKAFAKFNINTERTKLTAQAVNKLAELIM 1010
QY 1384 GYKILSEKYSKSLDSIKKIYNKQGENEKYLPFLANNIETLYTVNDKIDLYPIHLEAVY 1443


```

Db 1011 NRKDEKIDRKAKNTODLKK-----KEKENRK-----1040
Qy 1444 LNTYREKSNVAVKIKELNYLTKTQDKLADFKKNNNFVCIADLSTUDYNNNNLTKL-----1499
Db 1041 LNRERKEN-OMVYKHOKELMDMOALVEECAHREIOMOLASKSDIEOLRAKLDDSD 1099
Qy 1500 STGWFEIARLAKTVLSNLDGNLQGMINTISOHCYKQCPONSGCFRHLDEREE-----CK 1554
Db 1100 STVASPESADETGDNPESEIEGWLAV-----PRNGIKRIGKMKQYVAVSK 1148
Qy 1555 CLNATKQEGDKCVENPN 1571
Db 1149 KILFYNDQDKREOSNPS 1165

RESULT 57
US-08-755-587-185
; Sequence 185, Application US/087555587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRC2A cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 185:
; SEQUENCE CHARACTERISTICS
; LENGTH: 1535 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-755-587-185

```

Query Match 3.28; Score 272; DB 3; Length 1535;
 Best Local Similarity 19.28; Pred. No. 1.5e-06;
 Matches 255; Conservative 223; Mismatches 520; Indels 344; Gaps 62;

```

Qy 267 KNAKREKKRLYOAOYDLSTYKQLEAHNLISVLEKRIPTLKKNNIKELDKINEIK 326
Db 15 KEACNKKEXQLFIT-----PEADSLSPLOEQYENDKRSKRVSDIKKEVLAA 63
Qy 327 NPPANSCTPNTLLDKKK-IEEHEKE-----IKEIAKTIKFI-----DSL 368

```

```

Db 64 HHPVQSAEYSDTDFOSOKSPLYDHENASTLLPTPSKDVLSNLMISGKKSKYMSDEL 123
Qy 369 FTDPLELEYLRK-----KNIDISAKVETKESTE-----FNEYNGV--TYPLSYNDINNA 420
Db 124 KGNNTSEDFELKNIPMEKNDOVCALSENSKRVELLPEPKYIRGASPSKRVQFNOTNLRV 183
Qy 421 NELNSFGDLINFEYTKREPSKNIYTDNERKFFINEIKK-----IKIEKKIESD-----470
Db 184 IQKODEETLLPKITVHSDSEELFSDNENNNPVQVANERNRNALONTMELHETDILCYNE 243
Qy 471 ---KKS-----YEDRSKSLNDITKEYEKLNIYD-SKFNNNI-----D 505
Db 244 PIFKNSITVLYADIDGKOATQVSIKKDLS--SNIIYDLAEKKNSKOKHLKTLGODLPD 302
Qy 506 LTFNEKMMGK-----RYSYKVEKLTHHTFASYEN-----SKHLEKTLAKLTYME 551
Db 303 ILINDKIPDKIDDCKMAGPLDPISNHSPGSGFPTASNKELIKSEHNKSKMFFKDE 362
Qy 552 DY-----SLRNIYVEKELKYKNLSKIENEIETLV-----ENIKKDEQLFEKK- 596
Db 363 EQYPTSLACVELVNTLALDNQKLSKPOLINTYSAHLQSSVYVSDCKNSHITPQMSFSKOD 422
Qy 597 ITKDEN-KPDEK--ILEVSDIVKVOYVYLLMNKIDELKTKQILKVELKHNHVP---650
Db 423 FNSNHNLTPOKAEITELSTLIEBSGOF---EFTQFRKPSYIQ---KMTFVPEPNO 474
Qy 651 ---NSTYKQENK-----QEPYILIVLKEIKDKLVMPYVESLINEKKNIKTEGSD 699
Db 475 VTILNTTEENDAGLVIMNAPSIOVNSSKQEGTVGIKQFAGLKDCKNSAGYLLD 534
Qy 700 NSEPTGEELTG-----OATTKPGQOA---GSALGDSVOAOQROKOAOPVPVYVPEA 751
Db 535 ENEVEFRGFYSAHGKVLNVTALQKAVKLSDIENISEKTSABVD-----PISLSS 587
Qy 752 KAQVPTPAPVNNKTENVSKLD-YLEKLYEFLNTSYICHKYLIVSHSTMNKILKQYVI- 809
Db 588 K-----FHDSVYSMFKIENHNDKTSKKCOLMLONNIEMTGTGFVEIETENYKIN 638
Qy 810 TKREESKLSSC-----DPLDILFNIONNI-----PYMYSM 839
Db 639 TENEDNKYTAASRNSRNLEFVGSDSSKNDYVCIHKDEKDLPTDORNICLKLSGOLMEG 698
Qy 840 FDSLNNLSLQFMELIYERKEMVCLYKLDNDKIKMLLEAKVSTSVK-----TLSS 892
Db 699 NTQIKEGSLDLFLVYKQOERCH-----GNMSNKKQL-TARKTBONIKDFEFDLSFQTA 752
Qy 893 SMQPLSLTPD-----KPEVSANDDTSHSTNLNLSKLFENILSLGKNNIYOEL 942
Db 753 SGKNIYVTKESFNKIVNFFDPKPE-----ELHNSLSNLSKLH---SDIRKNK---MDI 798
Qy 943 IQKSSSENFYKILKDSOTFYNESTFNFKSKADDINSLNDESKKKLEEDINKLKTILQ 1002
Db 799 LSHETDTVKKKILKESPVGTGNOLVTF-----QEPQGDDEI---KEPTL 841
Qy 1003 LSPDLYNNKYKILK-----ERLPDKK-----KTVGYKMQIKKLTLLK 1039
Db 842 LGFHTASGKKYKIKTESLDYKKNLFDEKEQGETSETTSHOWAKTL-KYREACQDLIELAC 900
Qy 1040 EQLE-----SKLNSLNPRKHLQNFVFPNKKKEALIEAETENTLTKLKHNYGL 1091
Db 901 ETVETITAPCKCKEMQNSLNNDKIVCIETVVPKLLSDMLCTQTEMLTKSKSIFLK-----955
Qy 1092 VKYVNGESSPLKLTSEESIQEDNYSLEN-----FKVLSKLEGLKQDLNLEKKLSTYS 1147
Db 956 KVHVENVEKETAKSPTTCYTQSPSYIENSALALAYTSCSRCSVSQTSLEAKO--WLR 1013
Qy 1148 SGLHLIALELKEVINKNYNTSPSENNTDVNNALESYKFKPLPBGTVATVYVSESGDPL 1207
Db 1014 EG---IFPDQPERINTADYVGNLSYENNSK-NTIAESKNH-----PSEKQDPTL 1059
Qy 1208 EOSQPKKPASTHVGAESFTTTSQNVDEVDVVIIVPIFGSEEDYDGLGOVYGEAVTP 1267
Db 1060 NNS-----SMSNSYSYL-----SDEVYSDSGYLSKKNK--LD 1088

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 22:00:50 ; Search time 9477 Seconds

(without alignments)
15170.187 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940

Sequence: 1 cgcacgcgcatgaataatcat.....ttcatctaatagatcgtg 4940

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 180 summaries

Database :

GenEmbl:.*
1: gb_da.*
2: gb_bg.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB ID	Description
1	4940	100.0	4940	6 A92451	A92451 Sequence 2
2	4938.4	100.0	4940	12 CVE131294	AJ131294 Cloning v
3	2816.4	57.0	5917	3 PFAMF195	X02919 Plasmodium
4	2816.4	57.0	7038	3 PEGP195A	X15063 Plasmodium
5	2814.4	57.0	4920	6 A92450	A92450 Sequence 1
6	2706.2	54.8	5760	6 A04562	A04562 P.falcipar
7	2706.2	54.8	5760	6 E00656	E00656 cDNA encodi
8	2688.8	54.4	5282	3 PEP180	X03371 P.falciparu
9	1400.6	28.4	5271	3 AF062349	AF062349 Plasmodu
10	1331.2	26.9	5243	3 AF062348	AF062348 Plasmodu
11	1330.8	26.9	5085	3 AF218248	AF218248 Plasmodu
12	1290.4	26.1	5392	3 PEP190G1	X05624 Plasmodium
13	1290.4	26.1	5754	3 PFAMMSAA	M9143 P.falciparu
14	1272.2	25.6	5312	3 PFMEZSA1B	Z53327 P.falciparu
15	1267.2	25.7	5276	3 PFAGP195	M37213 P.falciparu
16	1267.2	25.7	5541	3 PEGP195	X03831 P.falciparu
17	1169.6	23.7	1968	3 AF480451	AF480451 Plasmodu
18	910.4	18.4	1636	3 PFAMSP806	D13357 P.falcipar
19	910.4	18.4	1636	3 PFAMSP808	D13358 P.falcipar
20	908.8	18.4	1636	3 PFAMSP828	D13360 P.falcipar
21	908.8	18.4	1636	3 PFAMSP837	D13362 P.falcipar
22	908.8	18.4	1636	3 PFAMSP844	D13363 P.falcipar
23	907.2	18.4	1636	3 PFAMSP844	D13363 P.falcipar
24	907.2	18.4	1636	3 PFAMSP844	D13363 P.falcipar
25	905.6	18.3	1636	3 PFAMSP8222	D13359 P.falcipar
26	784.6	15.9	3395	3 PFMSMSA1	X61930 P.falcipar
27	711.8	14.4	3518	3 PFAP190A	M35727 P.falciparu
28	711.8	14.4	3518	3 PEP190G	Y00087 Plasmodium
29	691	14.0	1203	3 PFAMSP1	M4681 Plasmodium
30	655.8	13.3	1131	3 AF325919	AF325919 Plasmodu
31	623.6	12.6	1065	3 PFASURFPO	L20092 Plasmodu
32	570.2	11.5	999	3 PFMSA1AC	M77715 Plasmodium
33	570.2	11.5	999	3 PFMSA1AC	M77715 Plasmodium
34	570.2	11.5	999	3 PFMSA1AC	M77715 Plasmodium
35	568.6	11.5	999	3 PFMSA1AM	M77725 Plasmodium
36	568.6	11.5	999	3 PFMSA1AF	M77718 Plasmodium
37	556.6	11.3	1133	3 AF061143	AF061143 Plasmodu
38	547.8	11.1	1080	3 AF286876	AF286876 Plasmodu
39	546.4	11.1	1956	3 PFMEZSA1D	Z53329 P.falciparu
40	546.2	11.1	999	3 PFMSA1AH	M77720 Plasmodu
41	544.8	11.0	1956	3 PFMEZSA1A	Z53326 P.falciparu
42	540.4	10.9	1141	3 AF061141	AF061141 Plasmodu
43	538.8	10.9	1142	3 AF061140	AF061140 Plasmodu
44	535.4	10.8	1959	3 PFMEZSA1C	Z53328 P.falciparu
45	528.2	10.7	1116	3 AF061146	AF061146 Plasmodu
46	527.8	10.7	1150	3 PFHB3A	X52962 Plasmodu
47	527.2	10.4	1158	3 AF061136	AF061136 Plasmodu
48	512	10.4	1158	3 AF061147	AF061147 Plasmodu
49	508.2	10.3	1102	3 AF061135	AF061135 Plasmodu
50	508.2	10.3	1157	3 AF061145	AF061145 Plasmodu
51	502.4	10.2	1123	3 AF061144	AF061144 Plasmodu
52	500.8	10.1	1124	3 AF061144	AF061144 Plasmodu
53	500.2	10.1	1149	3 AF061138	AF061138 Plasmodu
54	499.6	10.1	1167	3 AF061137	AF061137 Plasmodu
55	498.2	10.1	1129	3 AF061133	AF061133 Plasmodu
56	496.4	10.0	1002	3 PFAMSA1AD	M77716 Plasmodu
57	496.4	10.0	1002	3 PFAMSA1AK	M77723 Plasmodu
58	485.2	9.8	1002	3 PFAMSA1AB	M77714 Plasmodu
59	485.2	9.8	1002	3 PFAMSA1AL	M77724 Plasmodu
60	484.6	9.8	1110	3 AF061132	AF061132 Plasmodu
61	483.2	9.8	1121	3 AF061142	AF061142 Plasmodu
62	483	9.8	1111	3 AF061134	AF061134 Plasmodu
63	483	9.8	1112	3 AF061131	AF061131 Plasmodu
64	473.2	9.6	993	3 PFAMSA1AP	M77728 Plasmodu
65	463.2	9.4	984	3 PFAMSA1AJ	M77722 Plasmodu

66	463.2	9.4	984	3	PFAMSAIAN	M77726 Plasmodium	139	246.6	5.0	354	6	A64813	A64813 Sequence 4
67	457	9.3	1102	3	AF061150	AF061150 Plasmodium	140	243.4	4.9	737	6	A01099	A01099 Artificial
68	455.4	9.2	1101	3	AF061149	AF061149 Plasmodium	141	240.2	4.9	668	6	E01429	E01429 Genomic DNA
69	454	9.2	1141	3	AF061120	AF061120 Plasmodium	142	237.2	4.8	639	3	PFAMSAIAN	M77735 Plasmodium
70	440.4	8.9	1133	3	AF061119	AF061119 Plasmodium	143	228.6	4.6	1774	3	AF199407	AF199407 Plasmodium
71	438.4	8.9	1177	3	AF061122	AF061122 Plasmodium	144	227	4.6	1774	3	AF199405	AF199405 Plasmodium
72	437.4	8.9	963	3	PFAMSAIAU	M77733 Plasmodium	145	217.6	4.4	668	6	I05209	I05209 Sequence 5
73	435.6	8.8	1726	3	PFAMSP8352	D13349 P. falcipar	146	214.2	4.3	371	3	AB016617	AB016617 Plasmodium
74	434	8.8	1726	3	PFAMSP815	D13346 P. falcipar	147	212.6	4.3	371	3	AB016616	AB016616 Plasmodium
75	434	8.8	1726	3	PFAMSP838	D13350 P. falcipar	148	212.6	4.3	371	3	AB016615	AB016615 Plasmodium
76	432.4	8.8	1726	3	PFAMSP807	D13344 P. falcipar	149	212	4.3	442	3	PP190B	PP190B Plasmodium
77	432.4	8.8	1726	3	PFAMSP8351	D13348 P. falcipar	150	210.4	4.3	365	3	D86292	D86292 Plasmodium
78	432.4	8.8	1726	3	PFAMSP842	D13352 P. falcipar	151	209.4	4.2	371	3	AB016618	AB016618 Plasmodium
79	430.8	8.7	1726	3	PFAMSP947	D13355 P. falcipar	152	207.8	4.2	371	3	D86293	D86293 Plasmodium
80	429.2	8.7	1726	3	PFAMSP946	D13354 P. falcipar	153	205.6	4.2	1795	3	AF199409	AF199409 Plasmodium
81	428.4	8.7	1114	3	AF061123	AF061123 Plasmodium	154	202.4	4.1	415	3	PFAMSAI	M32116 P. falciparu
82	427.8	8.7	1729	3	PFAMSP814	D13345 P. falcipar	155	201.8	4.1	342	6	A64801	A64801 Sequence 6
83	427.8	8.7	1729	3	PFAMSP8221	D13347 P. falcipar	156	201.8	4.1	342	6	A64815	A64815 Sequence 6
84	426	8.6	1726	3	PFAMSP841	D13351 P. falcipar	157	201.8	4.1	456	6	A02300	A02300 Synthetic D
85	426	8.6	1726	3	PFAMSP843	D13353 P. falcipar	158	201.8	4.1	456	6	A02299	A02299 Synthetic D
86	424.6	8.6	1729	3	PFAMSPM20	D13343 P. falcipar	159	201.6	4.1	1801	3	AF199406	AF199406 Plasmodium
87	422.8	8.6	1135	3	AF061125	AF061125 Plasmodium	160	201.4	4.1	350	3	D86296	D86296 Plasmodium
88	422	8.5	1090	3	PE3D7A	X52963 Plasmodium	161	200.8	4.1	1795	3	AF199410	AF199410 Plasmodium
89	420	8.5	1104	3	AF061151	AF061151 Plasmodium	162	199.6	4.0	1801	3	AF199393	AF199393 Plasmodium
90	420	8.5	1107	3	AF061148	AF061148 Plasmodium	163	195.2	4.0	1801	3	AF199401	AF199401 Plasmodium
91	418.2	8.5	1107	3	AF061126	AF061126 Plasmodium	164	194.8	3.9	291	6	A64810	A64810 Sequence 1
92	413.2	8.4	1002	3	PFAMSAIAS	M77731 Plasmodium	165	194.8	3.9	291	6	A64810	A64810 Sequence 1
93	413.2	8.4	1002	3	PFAMSAIAT	M77732 Plasmodium	166	194.8	3.9	330	6	A64804	A64804 Sequence 9
94	411.4	8.3	5730	3	PEAPV200A	M75674 Plasmodium	167	194.8	3.9	330	6	A64818	A64818 Sequence 9
95	410.6	8.3	1141	3	AF061129	AF061129 Plasmodium	168	194.8	3.9	387	6	A64802	A64802 Sequence 7
96	409	8.3	1173	3	AF061130	AF061130 Plasmodium	169	194.8	3.9	387	6	A64816	A64816 Sequence 7
97	406.8	8.2	1098	3	AF061121	AF061121 Plasmodium	170	194.6	3.9	309	6	AX278285	AX278285 Sequence
98	398	8.1	939	3	PFAMSAIAO	M77727 Plasmodium	171	194.6	3.9	3147	6	AX278286	AX278286 Sequence
99	395.6	8.0	975	3	PFAMSAIAR	M77736 Plasmodium	172	194.6	3.9	3147	6	AX278287	AX278287 Sequence
100	394.6	8.0	966	3	PFAMSAIAX	AF061124 Plasmodium	173	193.2	3.9	1801	3	AF199395	AF199395 Plasmodium
101	393.8	8.0	1203	3	AF061124	M77729 Plasmodium	174	191.6	3.9	1801	3	AF199402	AF199402 Plasmodium
102	392.2	7.9	1005	3	PFAMSAIAQ	M77729 Plasmodium	175	191.6	3.9	1801	3	AF199403	AF199403 Plasmodium
103	392.2	7.9	1005	3	PFAMSAIAY	M77734 Plasmodium	176	191.6	3.9	1801	3	AF199403	AF199403 Plasmodium
104	392	7.9	698	3	PFMSPIA	X75899 P. falciparu	177	191.6	3.9	309	6	AX278288	AX278288 Sequence
105	392	7.9	698	3	PFMSPIREC	X78207 P. falciparu	178	191.4	3.9	3147	6	AX278289	AX278289 Sequence
106	389.8	7.9	966	3	PFAMSAIAY	M77737 Plasmodium	179	191.4	3.9	3147	6	AX278289	AX278289 Sequence
107	386.4	7.8	1030	3	PFAGP185	M14632 Plasmodium	180	191.4	3.9	3147	6	AX278290	AX278290 Sequence
108	384.4	7.8	1030	3	AF061127	AF061127 Plasmodium							
109	382.4	7.7	5181	3	PFAMSAIG	M0807 Plasmodium							
110	380.8	7.7	657	3	PFMSAIFID	Y10598 P. falciparu							
111	373.8	7.6	1277	3	AF061128	AF061128 Plasmodium							
112	360.8	7.3	635	3	PFU09209	U09209 Plasmodium							
113	360	7.3	720	3	AF251345	AF251345 Plasmodium							
114	353.6	7.2	969	3	PFAMMFA	L10380 Plasmodium							
115	337.4	6.8	893	3	PFAMMSAB	M19144 P. falciparu							
116	335.4	6.8	1116	3	PFU20730	U20730 Plasmodium							
117	333.8	6.8	1116	3	AF040569	AF040569 Plasmodium							
118	333.8	6.8	1116	3	PFU20731	U20731 Plasmodium							
119	332.4	6.7	1119	3	PFU20729	U20729 Plasmodium							
120	332.2	6.7	1116	3	PFU20655	U20655 Plasmodium							
121	332.2	6.7	1116	3	PFU20656	U20656 Plasmodium							
122	330.8	6.7	1119	3	PFU20727	U20727 Plasmodium							
123	330.8	6.7	1119	3	PFU20727	U20727 Plasmodium							
124	330.6	6.7	1116	3	PFU20653	U20653 Plasmodium							
125	329.2	6.7	1119	3	PFU20728	U20728 Plasmodium							
126	329.2	6.7	1119	3	PFU20732	U20732 Plasmodium							
127	327.6	6.6	1119	3	PFU20654	U20654 Plasmodium							
128	327.6	6.6	1119	3	AF040567	AF040567 Plasmodium							
129	327.6	6.6	1119	3	PFU20733	U20733 Plasmodium							
130	326	6.6	1119	3	AF040568	AF040568 Plasmodium							
131	325.4	6.6	666	3	PFAMSAIAT	M77721 Plasmodium							
132	322.8	6.5	1249	12	AF395808	AF395808 Synthetic							
133	308.8	6.3	660	6	A02298	A02298 Synthetic D							
134	308.8	6.3	1068	6	A02301	A02301 Synthetic D							
135	282.6	5.7	639	3	PFAMSAIAE	M77717 Plasmodium							
136	272	5.5	654	3	PFMSAIFIS	Y10599 P. falciparu							
137	268.8	5.4	654	3	PFMSAIFIS	Y10600 P. falciparu							
138	246.6	5.0	354	6	A64799	A64799 Sequence 4							

ALIGNMENTS

RESULT 1
LOCUS A92451 4940 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9814583.
ACCESSION A92451
VERSION A92451.1 GI:6741181

KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 4940)
AUTHORS Pan, W. and Bujard, H.
TITLE METHOD FOR PRODUCING RECOMBINANTS INTENDED FOR USE IN A COMPLETE
JOURNAL MALARIA ANTIGENE (p190)/MSPI
PATENT: WO 9814583-A 2 09-APR-1998;
PAN HERMANN (DE)
FEATURES
source 1..4940
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 1669 a 1157 c 1060 g 1054 t
ORIGIN
Query Match 100.0%; Score 4940; DB 6; Length 4940;
Best Local Similarity 100.0%; Pred No. 0;
Matches 4940; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	CGACGGGTATGAAATCATTTTCTTCCTCTGTCATTTCTGTATTCATCAATACT	60
Db	1	CGACGGGTATGAAATCATTTTCTTCCTCTGTCATTTCTGTATTCATCAATACT	60
Oy	61	CAGAGCTGACCCACGAATCTTATCAGAGACTGTTAAGAACTGGAACTTGGAAAT	120
Db	61	CAGAGCTGACCCACGAATCTTATCAGAGACTGTTAAGAACTGGAACTTGGAAAT	120
Oy	121	GCCGTCCTTACCCGATACAGCCTGTTCAGAAAGAGAGATGGTCTGATATGAAAGACG	180
Db	121	GCCGTCCTTACCCGATACAGCCTGTTCAGAAAGAGAGATGGTCTGATATGAAAGACG	180
Oy	181	AGTGGACAGGCCCTTACAACACACACCCCGTTCTAAAGGGTCTGTGGTACGGTGGC	240
Db	181	AGTGGACAGGCCCTTACAACACACCCCGTTCTAAAGGGTCTGTGGTACGGTGGC	240
Oy	241	TCCCGTGGGTCTGTGGGCTGTGGGGGTCCGTCGCCCTCCGGGGGACGGTGGCATCAGGT	300
Db	241	TCCCGTGGGTCTGTGGGCTGTGGGGGTCCGTCGCCCTCCGGGGGACGGTGGCATCAGGT	300
Oy	301	GGCTCAGTGGCAAGCGCGGCTTCGGGGAACAGTCGAAGAACCAATCTCATCACAATCT	360
Db	301	GGCTCAGTGGCAAGCGCGGCTTCGGGGAACAGTCGAAGAACCAATCTCATCACAATCT	360
Oy	361	AGCGATTCCGACGCGCAAGTCTTACGCCGACCTTACACACCGAGTGAAGAACTATCTCTC	420
Db	361	AGCGATTCCGACGCGCAAGTCTTACGCCGACCTTACACACCGAGTGAAGAACTATCTCTC	420
Oy	421	ACTATCAGGAGCTGAAGTACCCACAGTTGTCGACCTCATATCATATGCTGACACTG	480
Db	421	ACTATCAGGAGCTGAAGTACCCACAGTTGTCGACCTCATATCATATGCTGACACTG	480
Oy	481	TGTGATTAACATTCATGCGCTTCAAAATCTGATTGACGTTTACGAAGAGATCAATGACTC	540
Db	481	TGTGATTAACATTCATGCGCTTCAAAATCTGATTGACGTTTACGAAGAGATCAATGACTC	540
Oy	541	CTGATCAAGTTGAATTTCTCTCGACTTCGACTTCGTAAGGGCCAAACTGAATGACCTTTGGCC	600
Db	541	CTGATCAAGTTGAATTTCTCTCGACTTCGACTTCGTAAGGGCCAAACTGAATGACCTTTGGCC	600
Oy	601	AATGACTATTGTCCAAATTCATTCATTTGAAGATAGAGCCACAGAGTTGGACGTATTG	660
Db	601	AATGACTATTGTCCAAATTCATTCATTTGAAGATAGAGCCACAGAGTTGGACGTATTG	660
Oy	661	AAGAAATTGGTCTTCGGATATTCGCAAGCCTCTCGACACATCATAGACATGTGGGAAG	720
Db	661	AAGAAATTGGTCTTCGGATATTCGCAAGCCTCTCGACACATCATAGACATGTGGGAAG	720
Oy	721	ATGGAAGATTATTTAAAGAAATAGAAAGACCATGSAACATTAAAGAGCTGATGCAA	780
Db	721	ATGGAAGATTATTTAAAGAAATAGAAAGACCATGSAACATTAAAGAGCTGATGCAA	780
Oy	781	GAATCAAAAAAGACCATAGACAAATAATAGAAATGCAACCAAGAGAGAAAAAGAAAG	840
Db	781	GAATCAAAAAAGACCATAGACAAATAATAGAAATGCAACCAAGAGAGAAAAAGAAAG	840
Oy	841	TTGTACCAAGGCCCACTACGACTGTCTCATCTATACAAACAGCTTGAAGAGCCCATAC	900
Db	841	TTGTACCAAGGCCCACTACGACTGTCTCATCTATACAAACAGCTTGAAGAGCCCATAC	900
Oy	901	CTCATACGCGTACTGGGAAGGCCCATAGACACCTCATAGAAATGAAAAATTTCAAAGAA	960
Db	901	CTCATACGCGTACTGGGAAGGCCCATAGACACCTCATAGAAATGAAAAATTTCAAAGAA	960
Oy	961	CTGCTGCACAAGATTATGAAATTTAAGAAATCTCCGCCAGCCACTGTGGGAACCCCT	1020
Db	961	CTGCTGCACAAGATTATGAAATTTAAGAAATCTCCGCCAGCCACTGTGGGAACCCCT	1020
Oy	1021	AACACGCTGCTGGACAAAGAAAGAAATAGAGAGACAGAGAAAGATCAAGAGATC	1080
Db	1021	AACACGCTGCTGGACAAAGAAAGAAATAGAGAGACAGAGAAAGATCAAGAGATC	1080

QY	1081	GCMAAACCTAATAGTTCACACATATAGATTCCTCTTACTGATACCCCTGACTGAGATAC	1140
Db	1081	GCMAAACCTAATAGTTCACACATATAGATTCCTCTTACTGATACCCCTGACTGAGATAC	1140
QY	1141	TACTTGAGAGAGAATTAAGATTAATAGACATCTCCGCAAGTGAGACAAAGATCA	1200
Db	1141	TACTTGAGAGAGAATTAAGATTAATAGACATCTCCGCAAGTGAGACAAAGATCA	1200
QY	1201	ACCGAACCTAATGAATATCCCAATGGTGTGAGCTACCCCTGTCTTATPAAGATATCAAC	1260
Db	1201	ACCGAACCTAATGAATATCCCAATGGTGTGAGCTACCCCTGTCTTATPAAGATATCAAC	1260
QY	1261	AACGCTCTCAACGAGCTCAATFACCTTCGGTACTGATTAACCCCTTGATTAACGAAA	1320
Db	1261	AACGCTCTCAACGAGCTCAATFACCTTCGGTACTGATTAACCCCTTGATTAACGAAA	1320
QY	1321	GAACCTCTTAAGATTAATCTACACAGCAATGAGAGAAAGAAATTTATACGAAATTCAG	1380
Db	1321	GAACCTCTTAAGATTAATCTACACAGCAATGAGAGAAAGAAATTTATACGAAATTCAG	1380
QY	1381	GAGAAAGATCAAAATTTGAGAGAGAAAAATTTGAGAGTGCACAGAAAAGTTTACGAGACCGC	1440
Db	1381	GAGAAAGATCAAAATTTGAGAGAGAAAAATTTGAGAGTGCACAGAAAAGTTTACGAGACCGC	1440
QY	1441	AGCAAAAGTCTAAACGATATACCTAAAGAGTATGAAAAGCTGCTACAGAGATCTATGAT	1500
Db	1441	AGCAAAAGTCTAAACGATATACCTAAAGAGTATGAAAAGCTGCTACAGAGATCTATGAT	1500
QY	1501	TCCAAATTCACAACTATACCTGACGCTGACCACCTTCGCAAAAATGATGGGAAAACGGTAC	1560
Db	1501	TCCAAATTCACAACTATACCTGACGCTGACCACCTTCGCAAAAATGATGGGAAAACGGTAC	1560
QY	1561	TCTTACAAAGTGGAGAACTGACACACACCAATATACCTTTCGATCCATAGAGAAATCTTAG	1620
Db	1561	TCTTACAAAGTGGAGAACTGACACACACCAATATACCTTTCGATCCATAGAGAAATCTTAG	1620
QY	1621	CATAATCTTGAGAGCTCACCAAAAGCTCTTAAGTATATGAGAGCAATTCCTCGGGAAC	1680
Db	1621	CATAATCTTGAGAGCTCACCAAAAGCTCTTAAGTATATGAGAGCAATTCCTCGGGAAC	1680
QY	1681	ATTGTGTGGAGAAAGAACTAAAGTATTTACAGAAATCTCATAGATGACATGGAACGAG	1740
Db	1681	ATTGTGTGGAGAAAGAACTAAAGTATTTACAGAAATCTCATAGATGACATGGAACGAG	1740
QY	1741	ATCGAGAGCGCTGTGTGAGAACATTAAAGAGAGATGAGAAACATGTTTGTGAGAGAGATT	1800
Db	1741	ATCGAGAGCGCTGTGTGAGAACATTAAAGAGAGATGAGAAACATGTTTGTGAGAGAGATT	1800
QY	1801	ACAAAAGCGAAANTTAACACGATGAGAGAGATCTGAGAGTCTCCGATATTTGTTAAAGTC	1860
Db	1801	ACAAAAGCGAAANTTAACACGATGAGAGAGATCTGAGAGTCTCCGATATTTGTTAAAGTC	1860
QY	1861	CAAGTCAGAAAGTGTCCATCGATGAACCAAGATTGAATGAATCAACACATCACTATT	1920
Db	1861	CAAGTCAGAAAGTGTCCATCGATGAACCAAGATTGAATGAATCAACACATCACTATT	1920
QY	1921	CTGAAGAGACGTGGAGTTAAACATTAATATACATGTCCCAATAGTTATTAACAGAGCAT	1980
Db	1921	CTGAAGAGACGTGGAGTTAAACATTAATATACATGTCCCAATAGTTATTAACAGAGCAT	1980
QY	1981	AAGGAGAGACCATTAATCACTCATCTGATCTCAAGAAAAGAGATGAGCAAACTGAAAGTCTTC	2040
Db	1981	AAGGAGAGACCATTAATCACTCATCTGATCTCAAGAAAAGAGATGAGCAAACTGAAAGTCTTC	2040
QY	2041	ATGCCCCAAAGTCGAGAGCTCGATCAACGGAAGAGAGAGACATTTAAACTGAAAGGACAG	2100
Db	2041	ATGCCCCAAAGTCGAGAGCTCGATCAACGGAAGAGAGAGACATTTAAACTGAAAGGACAG	2100
QY	2101	TCAGATTAATCTCCGAGCTTCACAGAGAGAGAGATTAACCGGACAGGCTACCAACCAAGCC	2160
Db	2101	TCAGATTAATCTCCGAGCTTCACAGAGAGAGAGATTAACCGGACAGGCTACCAACCAAGCC	2160
QY	2161	GGACAAACAGGCGGTTTCAAGCTCTCGAAAGCGATAGCGTGCAAGCTCAAGCAAGAGACG	2220

|||||
Db 2161 GGACAAACAGCGCGTTCCACTCTCGAAGGCGATAGCGTCAAGCTCAACACAAAGAGCAG 2220
QY 2221 AAGCAGGACACACCTCCACTGCGCAGTGGCCGCTCCAGAGGCTTAAGCTCAAGTCCCTACA 2280
Db 2221 AAGCAGGACACACCTCCACTGCGCAGTGGCCGCTCCAGAGGCTTAAGCTCAAGTCCCTACA 2280
QY 2281 CCACACAGCTCCGTGTGAATAACAAGACCGAGAAATGTGAGCAAACTGAGTACCTTGAGAG 2340
Db 2281 CCACACAGCTCCGTGTGAATAACAAGACCGAGAAATGTGAGCAAACTGAGTACCTTGAGAG 2340
QY 2341 CTCATGAGTTCCTGAAATACATCTCAATCTGCGCACAATATATCTCTCTCAACG 2400
Db 2341 CTCATGAGTTCCTGAAATACATCTCAATCTGCGCACAATATATCTCTCTCAACG 2400
QY 2401 ACTATGAAAGAGAAATCTTAAAGCTACAAAGATACCAAGAGAGAGAGTAAAGT 2460
Db 2401 ACTATGAAAGAGAAATCTTAAAGCTACAAAGATACCAAGAGAGAGAGTAAAGT 2460
QY 2461 TCCCTCTGTGATCAGCTGAGCCTGCTGTCAATATCCAGAACAACTCCCGTTATGTAT 2520
Db 2461 TCCCTCTGTGATCAGCTGAGCCTGCTGTCAATATCCAGAACAACTCCCGTTATGTAT 2520
QY 2521 TCTATGTTGATAGCTTCAACAATTTCTCTCTCAACTGTTATGAGATATATGAGAG 2580
Db 2521 TCTATGTTGATAGCTTCAACAATTTCTCTCTCAACTGTTATGAGATATATGAGAG 2580
QY 2581 GAGATGCTGCAACCTGTATTAACCTCAAGAACAGACAGAAAGATTAGAACCCTTGAGAG 2640
Db 2581 GAGATGCTGCAACCTGTATTAACCTCAAGAACAGACAGAAAGATTAGAACCCTTGAGAG 2640
QY 2641 GAAGCTAAGAGGTCCTCAACCTCTGTGTTAAACTCTCTCTTCCAGCTCATGCAACCATG 2700
Db 2641 GAAGCTAAGAGGTCCTCAACCTCTGTGTTAAACTCTCTCTTCCAGCTCATGCAACCATG 2700
QY 2701 TCTCTCACACCTCAAGACAGCCGGAAGTGAAGGCTTAAGAGACACCTCTCACTGAGC 2760
Db 2701 TCTCTCACACCTCAAGACAGCCGGAAGTGAAGGCTTAAGAGACACCTCTCACTGAGC 2760
QY 2761 AACCTTAATTAACCTCACTGAAAGCTGTTGAGAACATCTCTCTCTCGGCAAGATTAAGAC 2820
Db 2761 AACCTTAATTAACCTCACTGAAAGCTGTTGAGAACATCTCTCTCTCGGCAAGATTAAGAC 2820
QY 2821 ATCTACCAAGAACTTATGAGACAAATGCTCGAGAACTCTTACGAGAAATACCTGAGAA 2880
Db 2821 ATCTACCAAGAACTTATGAGACAAATGCTCGAGAACTCTTACGAGAAATACCTGAGAA 2880
QY 2881 GACAGGACACATTTCTATACGAGAGCTTCACTAACTTGTGAAATCTTAAAGCCGATGAT 2940
Db 2881 GACAGGACACATTTCTATACGAGAGCTTCACTAACTTGTGAAATCTTAAAGCCGATGAT 2940
QY 2941 ATCAACTCTCTTAACGATGAATCTTAACGTAAGAACCTGGAAGAGACATCAATAAGCTG 3000
Db 2941 ATCAACTCTCTTAACGATGAATCTTAACGTAAGAACCTGGAAGAGACATCAATAAGCTG 3000
QY 3001 AAGAAAGACCTGCACTGAGCTTCCAGCTGTACACAACTGAACTGAACTGGAAGAGAA 3060
Db 3001 AAGAAAGACCTGCACTGAGCTTCCAGCTGTACACAACTGAACTGAACTGGAAGAGAA 3060
QY 3061 CTCCTGCAAGAAAGAGACAGTGGCAAGTATTAAGATCGATCAAGAGTGTGACTGTG 3120
Db 3061 CTCCTGCAAGAAAGAGACAGTGGCAAGTATTAAGATCGATCAAGAGTGTGACTGTG 3120
QY 3121 CTCGAAGAGACAGCTTGAAGACAACTCACTACTGAACATCCGAACAGACGTAAGCTGAG 3180
Db 3121 CTCGAAGAGACAGCTTGAAGACAACTCACTACTGAACATCCGAACAGACGTAAGCTGAG 3180
QY 3181 AACTTCTCAGTGTCTTCAACAAGAAAGAGAGAGCCGAGATGCGCGAGACAGAACTACT 3240
Db 3181 AACTTCTCAGTGTCTTCAACAAGAAAGAGAGAGCCGAGATGCGCGAGACAGAACTACT 3240
QY 3241 CTGGAGAAACCAAGATTTCTCTCAAACTACAAAGGCTGTCAAGTATTTATGTGCG 3300
Db 3241 CTGGAGAAACCAAGATTTCTCTCAAACTACAAAGGCTGTCAAGTATTTATGTGCG 3300

Db 3241 CTGGAGAAACCAAGATTTCTCTCAAACTACAAAGGCTGTCAAGTATTTATGTGCG 3300
QY 3301 GAGTCTTCTCTCTGAAAGCTCTCTCCGAGAGAGATCCAGACCGAGATTAATCTACGCC 3360
Db 3301 GAGTCTTCTCTCTGAAAGCTCTCTCCGAGAGAGATCCAGACCGAGATTAATCTACGCC 3360
QY 3361 AGCCTTGAGAACTTCAAGGCTCTGTCTAAGCTGGAAGGCAAGCTGAAGACAACTGAAAC 3420
Db 3361 AGCCTTGAGAACTTCAAGGCTCTGTCTAAGCTGGAAGGCAAGCTGAAGACAACTGAAAC 3420
QY 3421 CTGGAGAAAGAGAGTCAAGTCACTCTCTACGAGACCTCATACCTGATCCGCGAGCTC 3480
Db 3421 CTGGAGAAAGAGAGTCAAGTCACTCTCTACGAGACCTCATACCTGATCCGCGAGCTC 3480
QY 3481 AAGGAGTATTAAGAAACAAGAACTACACGGCAATAGGCCAAGCGAGATTAATACAGAC 3540
Db 3481 AAGGAGTATTAAGAAACAAGAACTACACGGCAATAGGCCAAGCGAGATTAATACAGAC 3540
QY 3541 GTGAATTAACGCACTGGAATCTTACAAAGAAATTCCTGCTGAAGGAACAGATGTCCGACT 3600
Db 3541 GTGAATTAACGCACTGGAATCTTACAAAGAAATTCCTGCTGAAGGAACAGATGTCCGACT 3600
QY 3601 GTGGTGTGGAATCTGGGCTCCGACACACGAGAGAGTCTCACTAAGAAAGCTGATCT 3660
Db 3601 GTGGTGTGGAATCTGGGCTCCGACACACGAGAGAGTCTCACTAAGAAAGCTGATCT 3660
QY 3661 ACTCATGTGAGCGCAGTGCATACATCAATACATCACTCAGAACCTCGACGATGAGTGC 3720
Db 3661 ACTCATGTGAGCGCAGTGCATACATCAATACATCACTCAGAACCTCGACGATGAGTGC 3720
QY 3721 GATGACGTATATTTGCTGCTATCTTCCGCGAGAGAGAGAGAGTACATGACCTCGGC 3780
Db 3721 GATGACGTATATTTGCTGCTATCTTCCGCGAGAGAGAGAGAGTACATGACCTCGGC 3780
QY 3781 GATGACGTATATTTGCTGCTATCTTCCGCGAGAGAGAGAGAGTACATGACCTCGGC 3840
Db 3781 GATGACGTATATTTGCTGCTATCTTCCGCGAGAGAGAGAGAGTACATGACCTCGGC 3840
QY 3841 GAGAAAGAAATAGCAAGTGTCTATCTGAAACCTTGCGAGGCGCTATAGTCTCTCAAG 3900
Db 3841 GAGAAAGAAATAGCAAGTGTCTATCTGAAACCTTGCGAGGCGCTATAGTCTCTCAAG 3900
QY 3901 AAAACAGCTGGAATATACCTGATGACCTTCAATGTCAAAGTGAAGGACATTCGAAACAC 3960
Db 3901 AAAACAGCTGGAATATACCTGATGACCTTCAATGTCAAAGTGAAGGACATTCGAAACAC 3960
QY 3961 CGCTTAATTAAGAGAAATTTCAAGAGAGTCTTGAGAGGCACTGATTCCTATAAA 4020
Db 3961 CGCTTAATTAAGAGAAATTTCAAGAGAGTCTTGAGAGGCACTGATTCCTATAAA 4020
QY 4021 GACCTGACCTCTCTAATACCTGTTGCAAGAGCCCATACAAATGCTCTCAATAAGAGAG 4080
Db 4021 GACCTGACCTCTCTAATACCTGTTGCAAGAGCCCATACAAATGCTCTCAATAAGAGAG 4080
QY 4081 AAGGATTAATTTCTGTCTGTTACAACTATATCAAGAGACCTCAAGCCGATTAAT 4140
Db 4081 AAGGATTAATTTCTGTCTGTTACAACTATATCAAGAGACCTCAAGCCGATTAAT 4140
QY 4141 TTGCGTAATGATGTGCTGGGATTTACAAAGATCTGAGGAAATATCAAGTCTGACTT 4200
Db 4141 TTGCGTAATGATGTGCTGGGATTTACAAAGATCTGAGGAAATATCAAGTCTGACTT 4200
QY 4201 GACTCTATTTAAAAAGTATATCAACGATTAAGCAAGGCGAGATGAATATCTGCCCTTC 4260
Db 4201 GACTCTATTTAAAAAGTATATCAACGATTAAGCAAGGCGAGATGAATATCTGCCCTTC 4260
QY 4261 CTGAATTAACATGGAACCTGTGACAAAGAGAGAGAGCAAAATATCAACCTCTCGTAAT 4320
Db 4261 CTGAATTAACATGGAACCTGTGACAAAGAGAGAGAGCAAAATATCAACCTCTCGTAAT 4320
QY 4321 CACCTGAGAGCCCAAGGCTCTCAACTTACTTACGAGAAAGCAATGTGAAGTTAAATC 4380
Db 4321 CACCTGAGAGCCCAAGGCTCTCAACTTACTTACGAGAAAGCAATGTGAAGTTAAATC 4380


```

QY 4381 AAGGAGCTGAAGTACCTCAAAAAACAATCCAGACAGCTGGAGATTTCAGAAAAATATAC 4440
Db 4381 AAGGAGCTGAAGTACCTCAAAAAACAATCCAGACAGCTGGAGATTTCAGAAAAATATAC 4440
QY 4441 AATTTCGTGGAAATTCGACAGCTGTCTACCGATTATATACCAACAATCTCTGACCAAG 4500
Db 4441 AATTTCGTGGAAATTCGACAGCTGTCTACCGATTATATACCAACAATCTCTGACCAAG 4500
QY 4501 TTTCTGTCCACTGGCAGTGGTGTTCGAAAAACCTGCCAAAACAGTGTGAGCAATCTGCTC 4560
Db 4501 TTTCTGTCCACTGGCAGTGGTGTTCGAAAAACCTGCCAAAACAGTGTGAGCAATCTGCTC 4560
QY 4561 GACGGCAACTGCAAGGGCATGTGAAATCTCCCGACACCAATGGCTGAAGAAACAGTGC 4620
Db 4561 GACGGCAACTGCAAGGGCATGTGAAATCTCCCGACACCAATGGCTGAAGAAACAGTGC 4620
QY 4621 CCCGAGAAATAGCGGCTGTTTCAGGCATCTGGACGACGCGGAGAGTGAAGTGTCTCCTG 4680
Db 4621 CCCGAGAAATAGCGGCTGTTTCAGGCATCTGGACGACGCGGAGAGTGAAGTGTCTCCTG 4680
QY 4681 AACTACAAACAAGAGAGATAGTGGCTGGAGAACCCAAACCTTACCTGCAATGAAAAC 4740
Db 4681 AACTACAAACAAGAGAGATAGTGGCTGGAGAACCCAAACCTTACCTGCAATGAAAAC 4740
QY 4741 AATGGCGGGTGTGACGGCGATGCTAAATGACACCGAGAGAGACAGGGCTTAACGGAAG 4800
Db 4741 AATGGCGGGTGTGACGGCGATGCTAAATGACACCGAGAGAGACAGGGCTTAACGGAAG 4800
QY 4801 AAAATCAGATCGAGTGTACTAAGCCGACATCTATCCACTCTTCGACGGGATTTTTCG 4860
Db 4801 AAAATCAGATCGAGTGTACTAAGCCGACATCTATCCACTCTTCGACGGGATTTTTCG 4860
QY 4861 TTCAGCTTAATTTCTCTGGGATCTCCCTTCCTGCTGATCTCTATCTGATCTGTACAGC 4920
Db 4861 TTCAGCTTAATTTCTCTGGGATCTCCCTTCCTGCTGATCTCTATCTGATCTGTACAGC 4920
QY 4921 TTCATCTAATAGATGATGAG 4940
Db 4921 TTCATCTAATAGATGATGAG 4940

RESULT 2
CVEI31294 4940 bp DNA Linear SYN 24-FEB-1999
LOCUS Cloning vector pBSK*-MSP-1s/FCB.
DEFINITION A1131294
ACCESSION A1131294
VERSION A1131294.1 GI:4455908
KEYWORDS Cloning vector; major surface protein-1; msp-1 gene.
SOURCE Cloning vector pBSK*-MSP-1s/FCB.
ORGANISM artificial sequences; vectors.
REFERENCE 1 (bases 1 to 4940)
AUTHORS Bujard,H.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1998) Bujard H., ZMBH - Zentrum fuer Molekulare
Biologie, Universitaet Heidelberg, Im Neuenheimer Feld 282, 69120
Heidelberg, BW, GERMANY
REFERENCE 2 (bases 1 to 4940)
AUTHORS Pan,W., Kavot,E., Tolle,R., Frank,R., Mosbach,R., Turbachova,I. and
Bujard,H.
TITLE Vaccine candidate MSP-1 from Plasmodium falciparum: a redesigned
4917 bp polynucleotide enables synthesis and isolation of
full-length protein from Escherichia coli and mammalian cells
JOURNAL Nucleic Acids Res. 27 (4), 1094-1103 (1999)
MEDLINE 99128299
PUBMED 9927744
FEATURES
source location/Qualifiers
1..4940
/organism="Cloning vector pBSK*-MSP-1s/FCB"
/db_xref="taxon:85834"
/focus
1..4940
source

```

```

/organism="Plasmodium falciparum"
/strain="FCB-1"
/db_xref="taxon:5833"
10..4929
/gene="msp-1"
10..4929
/gene="msp-1"
10..4929
/function="surface protein"
/codon_start=1
/evidence="experimental"
/product="major surface protein-1"
/protein_id="CAB36903.1"
/db_xref="GI:4455909"
/translacion="MKTIPLFCSSFLFIIMQCTHESYOELVKKLEALBDAVLTGS
LFOKEKAVLBGTSIGTAVTTSTPGSKSVASGSGGVSASGVSASGVSAS
GGSGNSRRTSPDSQSDAKSYADLKHVRNRYLLTKEKYPOLFDLTNMLTLCN
IHGFKYLIDGEEINELLYKLNFEFDLIRAKLNDVCANDYCOIIPFNKIRANLEDLVK
KLVGKRPDLNIDKNVGMEDYIKKKKTENINELIESKTIIDKNKATKEEEK
KLVOAYDLSTYNKLEAHNLISLEKRIIDTLKKNENIKELDKINIKNPPANG
NTPNLLDKNKKIIEHEKEIKETAKTIKPNIDSLFMDPLEVLYREKKNIDISARV
ETKSTEPNENPNCVYTPLSYNDINNALNELNSFGDILNPDYTKESKNIYDNEK
KFLNEIKETIKETKKIESDKSYEDSKSLNDITKEKELNLTIDSKNNNIDLTN
FEKMMGRYSYKVEKLNHTFASYESKSNLEKLTALYMEDYSLRNVYKELKY
YKNLISKENIEITLVENIKKDEOLEKKTIDENKPDKEILIESDILVVOYKVL
MKNIDELKTIOLILKVELKNIHVPSYKOENKOEYLYLVLEKEIDKLVEPKVE
SLINEEKNITKEGSDNSPSTGEITGQATTPRGQOASALGSDVOAOBOKA
QPPYVPVPEKKAQVTPPAPVNNKTENSLDYLEKLEPLNTSYLCHKYLIVSHS
MNEKTLIOYKITKEESKLSLSDPLDLPLFNQNNIPWYSFSDLSNLSOLEMEITE
KEWVCNLYKLDNDKIKNLEAKRVSTSVKTSSSSMQPLSDPPDKPEVSADOTS
HSYTNNSLKLFEINLSLGNKNIYQELIQKSSSENEKILKSDPFEVSFNPVK
SKADDISLNDSESKRLEEDINKLKTLOLSPDYKRYKLEKLEKLEKLVGKYK
OIKRTILKLEESKRLNSLNNPKHVLQNFVSFEPKKKEALTEALENTLEMKLILKHY
KGLIYYNNGESSPLKTLSESIOTREDNVASLENKYLSLEKLDKNDLNLEKRLSL
SSGLHLILAELEVKNNKNTGNSPESNNVDVNNALSKYKLEBEGDVTAVSESS
DTLESOQPKRPASTHVGAESNTITTSQNVDEVDVILVIFGSEEDYDLDGVVSG
EAVTPSYIDNLSKINENEYVLYLAKPAGVYRSRLKOLENNVMPFNVNVDILNSRN
KRENFKNVLESDDLIPYKDLTSSNVVDYPYFLKLEKREKFLSSYNIKDSIDTINF
ANDLYGYKILISEKYSKSDLSIKKYINDKQSENKYYPFLNNIETLYKTVNDKIDL
V
IHEAKVLYTYESKSNVEVKIKELNYLTIDKADRFKNNNPFVGLIDLSYVHNNL
LTKFLSTGMPFENLAKTVLSNLDGNLQGMNLISQHCYVKKQCCQNGSCRRHDERE
CKCLINTKQBGDCVENPNPFCNNNGGCDADACTEEDSGSKNKITTECTKPDSTP
LFDDIFCSSSNFIDIFPLILMLILYSFI"
10..69
/gene="msp-1"
sig_peptide
BASE COUNT 1669 a 1156 c 1060 g 1055 t
ORIGIN
Query Match 100.0%; Score 4938.4; DB 12; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CGCAGCGGTATGAANAATATATTTCTCTCTGTTTCATTTCTGTTTATATATCAATACT 60
Db 1 CGCAGCGGTATGAANAATATATTTCTCTCTGTTTCATTTCTGTTTATATCAATACT 60
QY 61 CAGTCCGTGACCCCGAATCTATCAGAGCTGGTTAAGAACTGGAAGAT 120
Db 61 CAGTCCGTGACCCCGAATCTATCAGAGCTGGTTAAGAACTGGAAGAT 120
QY 121 GCCGCTCTTACCGGATACAGCTGTTCAGAAAGGAGATGGTCAATGAAGGACG 180
Db 121 GCCGCTCTTACCGGATACAGCTGTTCAGAAAGGAGATGGTCAATGAAGGACG 180
QY 181 AGTGGCAGCGCGGTACACAGACACCGCGTTCTAAAGGGTGTGTGGCTAGGGTGGC 240
Db 181 AGTGGCAGCGCGGTACACAGACACCGCGTTCTAAAGGGTGTGTGGCTAGGGTGGC 240
QY 241 TCCGATGAGTCTGTGGGGGTCCGTCCCTCCGCGCAGCGAGCTGGCATCAGGT 300
Db 241 TCCGATGAGTCTGTGGGGGTCCGTCCCTCCGCGCAGCGAGCTGGCATCAGGT 300
QY 301 GGCTCAGTGGCAGCGGGGTTCCGGGAACAGTGAAGAACCAATCCATCTGACACTCT 360
Db 301 GGCTCAGTGGCAGCGGGGTTCCGGGAACAGTGAAGAACCAATCCATCTGACACTCT 360

```



```
Db 301 GGCCTAGTGGCAAGCGCGGTTCCGGGAACAGTCGAAGAACAATCCATCTGACAACTCT 360
QY 361 ACCGATTCGAGCCCAAGTCTACGCCGACCTCAACACACCGATGGAATATCTCTCTC 420
Db 361 ACCGATTCGAGCCCAAGTCTACGCCGACCTCAACACACCGATGGAATATCTCTCTC 420
QY 421 ACTATCAAGAGAGCTGAAGTACCCACAGTGTGTGACCTCATTATCATATGTGACACTG 480
Db 421 ACTATCAAGAGAGCTGAAGTACCCACAGTGTGTGACCTCATTATCATATGTGACACTG 480
QY 481 TGTGATTAACATTCATGCGCTTCAAAATATCGATTGAGGTTAGAAAGATTCATGAACTC 540
Db 481 TGTGATTAACATTCATGCGCTTCAAAATATCGATTGAGGTTAGAAAGATTCATGAACTC 540
QY 541 CTGTACAAGTTGAATTCTACTCTGCTGCTAAGGGCCAAACTGATGACGTTTCCGCC 600
Db 541 CTGTACAAGTTGAATTCTACTCTGCTGCTAAGGGCCAAACTGATGACGTTTCCGCC 600
QY 601 AATGACTATTGTCAAAATTCATTCATTTGAAGATCAGAGCCAAACGATTTGACGTATTG 660
Db 601 AATGACTATTGTCAAAATTCATTCATTTGAAGATCAGAGCCAAACGATTTGACGTATTG 660
QY 661 AAGAAGTGTCTTGGATATCGCAAGCCTCTCGACACATCAGACATGTGGCAAG 720
Db 661 AAGAAGTGTCTTGGATATCGCAAGCCTCTCGACACATCAGACATGTGGCAAG 720
QY 721 ATGGAAGATTATATTAAGAAATGAAGAACCATGAGAACATTTAGCGATCGAA 780
Db 721 ATGGAAGATTATATTAAGAAATGAAGAACCATGAGAACATTTAGCGATCGAA 780
QY 781 GAATCCAAAAGAACCATAGACAAAATAAGAAATGCACCAAGAGAGAAAAGAAAG 840
Db 781 GAATCCAAAAGAACCATAGACAAAATAAGAAATGCACCAAGAGAGAAAAGAAAG 840
QY 841 TTGTACCAAGCCCAAGTACCAAGTGTCCATCTATTAACAAAGCTTTAAAGAACCCCTAAC 900
Db 841 TTGTACCAAGCCCAAGTACCAAGTGTCCATCTATTAACAAAGCTTTAAAGAACCCCTAAC 900
QY 901 CTCATCAGCGTACTGGAAGAGCGCATAGACACCCCTCAAGAGAAATGAATATCAAGAA 960
Db 901 CTCATCAGCGTACTGGAAGAGCGCATAGACACCCCTCAAGAGAAATGAATATCAAGAA 960
QY 961 CTGCTCGACAAGATTATGAATTAAGAAATCTCCGCAAGCCAACTCTGGGAAACACCCCT 1020
Db 961 CTGCTCGACAAGATTATGAATTAAGAAATCTCCGCAAGCCAACTCTGGGAAACACCCCT 1020
QY 1021 AACAGCGTCTGGAAGAAACAAGAAATAGAGAGACAGAGAAAGATCAAAAGATC 1080
Db 1021 AACAGCGTCTGGAAGAAACAAGAAATAGAGAGACAGAGAAAGATCAAAAGATC 1080
QY 1081 GCCAAACCATTAAGTTCAACATAGATCTCTCTTACTGATCCCTTGAGCTGGAGTAC 1140
Db 1081 GCCAAACCATTAAGTTCAACATAGATCTCTCTTACTGATCCCTTGAGCTGGAGTAC 1140
QY 1141 TACTTGAGAGAGAAGATTAAGATATAGACATCTCCGCCAAAGTCGAGACAAAGGATCA 1200
Db 1141 TACTTGAGAGAGAAGATTAAGATATAGACATCTCCGCCAAAGTCGAGACAAAGGATCA 1200
QY 1201 ACCGAACCTTAATGAATCCCAATGCTGATACCCCTCTGCTTATTAACGATATCAAC 1260
Db 1201 ACCGAACCTTAATGAATCCCAATGCTGATACCCCTCTGCTTATTAACGATATCAAC 1260
QY 1261 AACGCTCTAAGAGAGCTCATATAGCTTGGTGAATTAACCCCTTGATTTACGAA 1320
Db 1261 AACGCTCTAAGAGAGCTCATATAGCTTGGTGAATTAACCCCTTGATTTACGAA 1320
QY 1321 GAACCTCTAAGAGATATCTACACAGCAATGAGAGAAAGAGTTTATCAACGAAATCAAG 1380
Db 1321 GAACCTCTAAGAGATATCTACACAGCAATGAGAGAAAGAGTTTATCAACGAAATCAAG 1380
QY 1381 GAGAGATCAAAATTTGAGAGAAAGAAATTTGAGAGTGAACAAAGATTACGAGACCGC 1440
Db 1381 GAGAGATCAAAATTTGAGAGAAAGAAATTTGAGAGTGAACAAAGATTACGAGACCGC 1440
QY 1441 AGCAAAAGTCTAAACGATATCACTAAAGATATGAAAAGCTGCTGACAGATCTATGAT 1500
Db 1441 AGCAAAAGTCTAAACGATATCACTAAAGATATGAAAAGCTGCTGACAGATCTATGAT 1500
QY 1501 TCCAAATTAACAAATTAACATTCGACCTGACCAACTTGAAGAAATGATGGGAAAACGCTAC 1560
Db 1501 TCCAAATTAACAAATTAACATTCGACCTGACCAACTTGAAGAAATGATGGGAAAACGCTAC 1560
QY 1561 TCTTACAAAGTGGAGAACTGACACACCATTAATACCTTTGCATCTCATGAGAAATTTAG 1620
Db 1561 TCTTACAAAGTGGAGAACTGACACACCATTAATACCTTTGCATCTCATGAGAAATTTAG 1620
QY 1621 CATTAATCTTGAGAGAGCTCAACAAAGCTTTAAGTATATGAGAGACTATTTCTGCGAAC 1680
Db 1621 CATTAATCTTGAGAGAGCTCAACAAAGCTTTAAGTATATGAGAGACTATTTCTGCGAAC 1680
QY 1681 ATTGTTGTGAGAGAAAGACTAAAGTATTAACAAGATCTCATTAAGTATCGAAAACGAG 1740
Db 1681 ATTGTTGTGAGAGAAAGACTAAAGTATTAACAAGATCTCATTAAGTATCGAAAACGAG 1740
QY 1741 ATCGAGACGCTTGTGAGAAACATTAAGAAAGATGAAGAACAGTTGTTGAGAAAGATTT 1800
Db 1741 ATCGAGACGCTTGTGAGAAACATTAAGAAAGATGAAGAACAGTTGTTGAGAAAGATTT 1800
QY 1801 ACAAAAGAGAAAATTAACACAGATGAGAAATCTGAGAGTCCGATATTTGTTAAAGTC 1860
Db 1801 ACAAAAGAGAAAATTAACACAGATGAGAAATCTGAGAGTCCGATATTTGTTAAAGTC 1860
QY 1861 CAAGTCAGAGAGTGTCTCTCATGAACAAGATTGAACTCAAGAACTCAACTCATTT 1920
Db 1861 CAAGTCAGAGAGTGTCTCTCATGAACAAGATTGAACTCAAGAACTCAACTCATTT 1920
QY 1921 CTGAAGAGAGTGGAGTTAAACATTAATATACATGTCCGAAATGATATAGCAGAGAAAT 1980
Db 1921 CTGAAGAGAGTGGAGTTAAACATTAATATACATGTCCGAAATGATATAGCAGAGAAAT 1980
QY 1981 AAGCAGGAACCATACTACTACTCTGATCTCAAGAAAAGATAGACAACTGAAAAGTGTTC 2040
Db 1981 AAGCAGGAACCATACTACTACTCTGATCTCAAGAAAAGATAGACAACTGAAAAGTGTTC 2040
QY 2041 ATGCCCAAAAGTGAAGAGCCTGATCAACGAAGAGAGAAAGAACTTAAACTGAAGACAG 2100
Db 2041 ATGCCCAAAAGTGAAGAGCCTGATCAACGAAGAGAGAAAGAACTTAAACTGAAGACAG 2100
QY 2101 TCAGATTAATCCGAGCCTTCCACAGAAAGAGAGATTAACCGGACGCTACACCAAGCCC 2160
Db 2101 TCAGATTAATCCGAGCCTTCCACAGAAAGAGAGATTAACCGGACGCTACACCAAGCCC 2160
QY 2161 GGACAACAGGCGGTTCCAGCTCTCGAAGGCGATAGCGTCAAGCTCAAGCACAAGAGCAG 2220
Db 2161 GGACAACAGGCGGTTCCAGCTCTCGAAGGCGATAGCGTCAAGCTCAAGCACAAGAGCAG 2220
QY 2221 AAGCAGGACAGCCTTCAAGTGCAGTGCAGTCCGCTTCCAGAGGCTTAACACTCAAGTCCCTA 2280
Db 2221 AAGCAGGACAGCCTTCAAGTGCAGTGCAGTCCGCTTCCAGAGGCTTAACACTCAAGTCCCTA 2280
QY 2281 CCACAGGCTCTCTGTAATTAACAAGACCGAGAAATGTAGCAAACTGAGACTTCTGGAAG 2340
Db 2281 CCACAGGCTCTCTGTAATTAACAAGACCGAGAAATGTAGCAAACTGAGACTTCTGGAAG 2340
QY 2341 CTCTATGAGTTCCTGAATACATCTACATCTGACACAAATATATCTCTCTCTCACAC 2400
Db 2341 CTCTATGAGTTCCTGAATACATCTACATCTGACACAAATATATCTCTCTCTCACAC 2400
QY 2401 ACTATGAAGAGAGAGATTTCTTAACAGTACAGATTAACCAAGAGAAAGAGAGTAAACAG 2460
Db 2401 ACTATGAAGAGAGAGATTTCTTAACAGTACAGATTAACCAAGAGAAAGAGAGTAAACAG 2460
QY 2461 TCCCTTGTGATCCACTGAGCCTGCTGCTCAATATCCAGAACACATCCCGTTATGTAT 2520
Db 2461 TCCCTTGTGATCCACTGAGCCTGCTGCTCAATATCCAGAACACATCCCGTTATGTAT 2520
```


QY 2521 TCTATGTTGATAGCCTCAACAATTCCTCTCTCACTGTTTCAGATATATAGAAG 2580
|||||
Db 2521 TCTATGTTGATAGCCTCAACAATTCCTCTCTCACTGTTTCAGATATATAGAAG 2580
QY 2581 GAGATGTTCTGCACCTGTATTAACCTCAAGACAAAGACAGATTAAGAACCTTGTGAG 2640
|||||
Db 2581 GAGATGTTCTGCACCTGTATTAACCTCAAGACAAAGACAGATTAAGAACCTTGTGAG 2640
QY 2641 GAAGTAAGAAGGTTCCACCTCTGTAAACCTCTCTCCAGCTCCATGCAACCACTG 2700
|||||
Db 2641 GAAGTAAGAAGGTTCCACCTCTGTAAACCTCTCTCCAGCTCCATGCAACCACTG 2700
QY 2701 TCTCTCACACCTCAAGACAGCCCAAGTGAAGCTTAAGACAGACCTCTCACTGAC 2760
|||||
Db 2701 TCTCTCACACCTCAAGACAGCCCAAGTGAAGCTTAAGACAGACCTCTCACTGAC 2760
QY 2761 AACCTTAATTAACCTCAAGACAGCTGTTTGAAGATCTCTCTCTGCGCAAGATTAAGAC 2820
|||||
Db 2761 AACCTTAATTAACCTCAAGACAGCTGTTTGAAGATCTCTCTCTGCGCAAGATTAAGAC 2820
QY 2821 ATCTACCAAGAACTTATGAGACAGAAATCGTCCGAACTTCTAGAGAGATTAAGAA 2880
|||||
Db 2821 ATCTACCAAGAACTTATGAGACAGAAATCGTCCGAACTTCTAGAGAGATTAAGAA 2880
QY 2881 GACAGCGACACATTTCTATTAAGAGAGCTTCACTAACTTCTGAAATCTAAAGCCGATGAT 2940
|||||
Db 2881 GACAGCGACACATTTCTATTAAGAGAGCTTCACTAACTTCTGAAATCTAAAGCCGATGAT 2940
QY 2941 ATCAACTCTTTAAAGATGAATCTTAAGCTTAAGAACTGGAAGAGACATCAATTAAGCTG 3000
|||||
Db 2941 ATCAACTCTTTAAAGATGAATCTTAAGCTTAAGAACTGGAAGAGACATCAATTAAGCTG 3000
QY 3001 AAGAAGACACTGCACTGAGCTTGCACCTGTACCAAGATCAAACTGAACTGGAGAGA 3060
|||||
Db 3001 AAGAAGACACTGCACTGAGCTTGCACCTGTACCAAGATCAAACTGAACTGGAGAGA 3060
QY 3061 CTCTTCGACAAAGAGAGACAGTGGCAATTAAGATGACATCAAGAACTTGAATGCTG 3120
|||||
Db 3061 CTCTTCGACAAAGAGAGACAGTGGCAATTAAGATGACATCAAGAACTTGAATGCTG 3120
QY 3121 CTCAAGGAGCAGCTTGAAGAGAACTCAACTGACGAACTCCGAAACACTACTGTCAG 3180
|||||
Db 3121 CTCAAGGAGCAGCTTGAAGAGAACTCAACTGACGAACTCCGAAACACTACTGTCAG 3180
QY 3181 AACTTCTCAGTGTCTTCAACAAGAGAGAGAGCCGAGATGCGCAGACAGAGAACT 3240
|||||
Db 3181 AACTTCTCAGTGTCTTCAACAAGAGAGAGAGCCGAGATGCGCAGACAGAGAACT 3240
QY 3241 CTGGAAGAACCCAGATTTCTCTCAACACTCAAAAGGCTCTGCAAGTATTATAAGGC 3300
|||||
Db 3241 CTGGAAGAACCCAGATTTCTCTCAACACTCAAAAGGCTCTGCAAGTATTATAAGGC 3300
QY 3301 GAGTCTTCTCTGGAAGACCTCTCCGAGAGAGACATCCAGAGATTAACATGCGC 3360
|||||
Db 3301 GAGTCTTCTCTGGAAGACCTCTCCGAGAGAGACATCCAGAGATTAACATGCGC 3360
QY 3361 AGCCTCGAAGACTTCAAGCTCTGCTCAAGCTCAAGGCAAGCTCAAGGACAACTGAAAC 3420
|||||
Db 3361 AGCCTCGAAGACTTCAAGCTCTGCTCAAGCTCAAGGCAAGCTCAAGGACAACTGAAAC 3420
QY 3421 CTGGAAGAGAGAGCTGAGTACCTCTCTAGCGAGCTGACATCACTGATGCGCGAGCTC 3480
|||||
Db 3421 CTGGAAGAGAGAGCTGAGTACCTCTCTAGCGAGCTGACATCACTGATGCGCGAGCTC 3480
QY 3481 AAGGAAGTCATTAAGAGAAAGAACTACACCGGCAATAGCCCAAGGAGAGATTAACAGAC 3540
|||||
Db 3481 AAGGAAGTCATTAAGAGAAAGAACTACACCGGCAATAGCCCAAGGAGAGATTAACAGAC 3540
QY 3541 GTGAATTAACGACCTGGAATCTTACAAAGATTCCTGCTGGAAGAGAAAGATGTCGCCACT 3600
|||||
Db 3541 GTGAATTAACGACCTGGAATCTTACAAAGATTCCTGCTGGAAGAGAAAGATGTCGCCACT 3600
QY 3601 GTGTGTCTGAATCTGCTCGACACACTGAGAGAGCTCTCAACCTTAAGAGCTGCAATCT 3660
|||||

Db 3601 GTGTGTCTGAATCTGCTCGACACACTGAGAGAGTCTCAACCTTAAGAGCTGCAATCT 3660
|||||
QY 3661 ACTCATGTCGAGCCGAGTCCAAATTAACATTAACACATCTCAGACGTCGACGATGAGTC 3720
|||||
Db 3661 ACTCATGTCGAGCCGAGTCCAAATTAACATTAACACATCTCAGACGTCGACGATGAGTC 3720
QY 3721 GATGAGCTCATATTTGCTCTTCCGCGAGAGAGAGAGAGATACGATACCTCGGC 3780
|||||
Db 3721 GATGAGCTCATATTTGCTCTTCCGCGAGAGAGAGAGATACGATACCTCGGC 3780
QY 3781 CAGGTGCTCACCGGAGGCTGTCACTCTCTCCGATTTGATTAACATTTCTGCCAAATC 3840
|||||
Db 3781 CAGGTGCTCACCGGAGGCTGTCACTCTCTCCGATTTGATTAACATTTCTGCCAAATC 3840
QY 3841 GAGAAAGCAATACGAAGTGTCTATCTGAACCTCTGCGAGGCTGTATAGTCTCAAG 3900
|||||
Db 3841 GAGAAAGCAATACGAAGTGTCTATCTGAACCTCTGCGAGGCTGTATAGTCTCTCAAG 3900
QY 3901 AAACAGCTGAGAAATACGATGATGACCTTCAATGTCACAGTGAAGACATTTGAAACAGC 3960
|||||
Db 3901 AAACAGCTGAGAAATACGATGATGACCTTCAATGTCACAGTGAAGACATTTGAAACAGC 3960
QY 3961 CGCTTTAATTAAGAGAGAAATTTCAAGACGCTTGGAGAGGACTGATTCCTATAAA 4020
|||||
Db 3961 CGCTTTAATTAAGAGAGAAATTTCAAGACGCTTGGAGAGGACTGATTCCTATAAA 4020
QY 4021 GACCTGACCTCCTCTTACTAGCTGTGTCAAGAGCCCATACAACTTCTCAATTAAGAGAG 4080
|||||
Db 4021 GACCTGACCTCCTCTTACTAGCTGTGTCAAGAGCCCATACAACTTCTCAATTAAGAGAG 4080
QY 4081 AAGGATAAATTTCTGTCTATGTTACAACTATATCAAGAGATCTCATGACACCGATATCAAT 4140
|||||
Db 4081 AAGGATAAATTTCTGTCTATGTTACAACTATATCAAGAGATCTCATGACACCGATATCAAT 4140
QY 4141 TTGCTAATGATGTCTGGGATTTTACAGAGATCTGAGCGAGAAATTAACAAGTCTGACCTT 4200
|||||
Db 4141 TTGCTAATGATGTCTGGGATTTTACAGAGATCTGAGCGAGAAATTAACAAGTCTGACCTT 4200
QY 4201 GACCTAATTAAGAAATTAATATCAAGATTAAGAGGCGAGAAATTAATGCTGCTTC 4260
|||||
Db 4201 GACCTAATTAAGAAATTAATATCAAGATTAAGAGGCGAGAAATTAATGCTGCTTC 4260
QY 4261 CTGAATTAACATGAAGAACCTCTGACAAAGACAGTGAACGCAAAATGCACTTCTGTAAT 4320
|||||
Db 4261 CTGAATTAACATGAAGAACCTCTGACAAAGACAGTGAACGCAAAATGCACTTCTGTAAT 4320
QY 4321 CACCTGAGGCGCAAGGCTCTCACTAATTAACGAGAGAGCAATGTGAAGTAAATC 4380
|||||
Db 4321 CACCTGAGGCGCAAGGCTCTCACTAATTAACGAGAGAGCAATGTGAAGTAAATC 4380
QY 4381 AAGGAGCTGAACTTACTCAAAACATCAAGCAAGCTGGGAGATTTCAAGAAAAATTAAC 4440
|||||
Db 4381 AAGGAGCTGAACTTACTCAAAACATCAAGCAAGCTGGGAGATTTCAAGAAAAATTAAC 4440
QY 4441 AATTCTGTCGGAATTAAGACCTGTCTACCGATTTAATCAACAAATCTCTGACAAAG 4500
|||||
Db 4441 AATTCTGTCGGAATTAAGACCTGTCTACCGATTTAATCAACAAATCTCTGACAAAG 4500
QY 4501 TTTCTGTCACACTGAGATGTTGAAAAACCTGCGCAAAACAGTCTGAGCAATCTGCTC 4560
|||||
Db 4501 TTTCTGTCACACTGAGATGTTGAAAAACCTGCGCAAAACAGTCTGAGCAATCTGCTC 4560
QY 4561 GACGCGCAACCTGCAAGGAGATGCTGAACATCTCCAGCAACATGAGTGAAGAAACAGTGC 4620
|||||
Db 4561 GACGCGCAACCTGCAAGGAGATGCTGAACATCTCCAGCAACATGAGTGAAGAAACAGTGC 4620
QY 4621 CCCCAGAAATAGGCGCTGTTTCAAGCATCTGGAAGAGGCGGAGAGATGAAGTGTCTCTG 4680
|||||
Db 4621 CCCCAGAAATAGGCGCTGTTTCAAGCATCTGGAAGAGGCGGAGAGATGAAGTGTCTCTG 4680
QY 4681 AACTCAAAACAAGAGAGATTAAGTGTGAGAGAACCCAAACCTTACCTGCAATGAAAC 4740
|||||

Db	4661	AACATCAACAAGAGAGATAGTACGCTGGAGAACCCAAACCCCTACTCGCAATGAAAC	4740
Qy	4741	AATGGCGGGGTGTGACGCGGATGCTAAATGACACCGAGAAACAGCGGCTCTAACGGAAG	4800
Db	4741	AATGGCGGGGTGTGACGCGGATGCTAAATGACACCGAGAAACAGCGGCTCTAACGGAAG	4800
Qy	4801	AAAATCACATGCGATGACTAGTACGACCGACACCTCTTCCACCTCTGACGGGATTTTGGC	4860
Db	4801	AAAATCACATGCGATGACTAGTACGACCGACACCTCTTCCACCTCTGACGGGATTTTGGC	4860
Qy	4861	TCCACCTCTTAAATTTCCCTGGGCACTCTTCCGCTGATCTCATCTGATCCTGTAACG	4920
Db	4861	TCCACCTCTTAAATTTCCCTGGGCACTCTTCCGCTGATCTCATCTGATCCTGTAACG	4920
Qy	4921	TTTCATCTAATAGATCGATG 4940	
Db	4921	TTTCATCTAATAGATCGATG 4940	
RESULT 3			
LOCUS	PEPNT195	5917 bp	mRNA linear INV 30-MAR-1995
DEFINITION	Plasmodium falciparum mRNA for major surface antigens precursor (P195).		
ACCESSION	X02919		
VERSION	X02919.1	GI:9864	
KEYWORDS	antigen: direct repeat; signal peptide; surface antigen.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum.		
REFERENCE	Enkaryota, Alveolata: Apicomplexa: Haemosporida: Plasmodium.		
AUTHORS	1 (bases 1 to 5917)		
	Holder, A.A., Lockyer, M.J., Odink, K.G., Sandhu, J.S.,		
	Riveros-Moreno, V., Nicholas, S.C., Hillman, Y., Davey, L.S.,		
	Tikarad, M.L.V., Schwarz, R.T. and Freeman, R.R.		
	Primary structure of the precursor to the three major surface		
	antigens of Plasmodium falciparum merozoites		
	Nature 317 (6034), 270-273 (1985)		
TITLE			
JOURNAL			
MEDLINE	86014355		
PUBMED	2995820		
REFERENCE	2 (bases 1 to 5917)		
AUTHORS	Holder, A.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-MAR-1991) Holder A.A., National Institute for Medical		
REMARK	Research, The Ridgeway, Mill Hill, London NW7 1AA, UK		
COMMENT	Update of published sequence		
FEATURES	Data kindly reviewed (06-MAR-1986) by R.R. Freeman.		
source	Location/Qualifiers		
	1..5917		
	/organism="Plasmodium falciparum"		
	/db_xref="taxon:5833"		
	418..5337		
	/note="P195 precursor"		
	/codon_start=1		
	/protein_id="CAA26676.1"		
	/db_xref="GI:9865"		
	/db_xref="SWISS-PROT:P04933"		
	/translation="MKITFEKSLPFLINOCYTHESGYLAKTKLADLVATGYS		
	LPQEKRYNLNCTGTATVTTSPSGSGSVASGSGSVASGSGSVASGSGSVAS		
	GGSSSRRTNPSDSDSAKSYADVLKRVNLYLTIKELYLPOLFDITNMLTCLDVS		
	IHGRKYLLDGEYELNELLYLNFYDLRALINVCADYQOIFPNLIRANEDLVYK		
	KLIVGEKRPDLINQDNVGMEDYIKKNKTKTEIDENKDEKILVEISDIKVOQVLT		
	KLYOAVDLSLYNKLQEAHNLIVLEKRIOTLKNENIKELDLINIKINPANS		
	NTPLIDLNKKIKIEHEKEIKETIAKTIKFNDLSFTPELELYLRKNKIDISAKYK		
	ETKSTEPNEIPNGVTTPILSTINDINNALNELNSGDLINPDIYKESKNITYDNERK		
	KFINIEIKRIKIERKKLESDDKSYEDRSKLNLDITKEYLEKLNEYDSKFNNNIDLTN		
	PEKMGKRSYKVEKLTNHNTPASYSKSNHLEKLTAKLYMEDYSLNINIVEKLEK		
	YKNLSIKLENIETLVENIKKDEQLEKTKTDENKDEKILVEISDIKVOQVLT		
	MNKIDELKTKQLILKNVGLKHNHVPNSYKQENKOEPLYLLVLEKIDKLVMPKVE		
	SLINEEKNTTBEOSNSPSTGELTNGQATKPRGOAGSALSGDSYQAOAOKOKQA		
	OPPPVPPPEAKQVPTTPPAPYNNKTENVSLDLYIEKTFELNTSYLCHKIIVLSH		
	MNEKLYKQYITKEEESKLSGCDPLDLFNLQNNIPWYSFDSLNNSLSQLEWETIS		
	KEWNCVNIKLDNDKIKNLLEAKVATSVYTLSSSQPLSTIPQDKPEVSANDDTIS		
	HSTUNNLSIKLENIILSGKKNKNIYQELIGKSSSENEFKILKSDIFYNSSFNPFVY		

[illegible]

Oy	370	GAGCCCAAGTCCTACGCCGACCTCAGACACCCAGAGAGAAACTATCTCCTCAGTATCAAG	429
Db	778	GATGCTAAATCTTACCGCTGATTTTAAACACAGAGTACGAAATTAAGTTACTTGTATCATATCAAA	837
Oy	430	GAGCTGAGTACCCACAGTGTTCGACCTCCACTCAATCATATCTGACACTGTGGATATAC	489
Db	838	GAACTCAAAATATCCTCAACCTTTTGATTTTAACTAAATCATATCTTAACCTTTGTGTGATAT	897
Oy	490	ATTATGCGCTTCAAAATATCTGATTTGACGGTTACGAAGATCATGAACCTCCTGTACAG	549
Db	898	ATTATGCTTCAAAATTTTAAATTTGATGATATGAAGAAATTAAGAAATTTATATATAA	957
Oy	550	TTGAAATTTACTTCGACTGCTACAGGGCAACGAATGAGCTTTGGGCCAATGAGCAT	609
Db	958	TTAAACTTTATTTTGATTTTATTAAGACAAAATTAATGATGTATGTGCTAATGATTTAT	1017
Oy	610	TGTCAAAATTCATTCATTTTGAAGATCAGACCCAAACGAGTTGGAGCTATTAAGAAGTTG	669
Db	1018	TGTCAAAATTCCTTTCATCTTAAATTCGTGCAAAATGAAATTTAGACGTACTTAAAAAACTT	1077
Oy	670	GCTCTGGATATGTGCAAGCCTCTCGACACATCAAGAGCAATATGTGGAAACATGGAAGAT	729
Db	1078	GTTTGGATATGAAAAACCATTAGCAATTTTAAAGATTAATGTAGAAAAATGGAAGAT	1137
Oy	730	TATATTTAAAAAGATTAAGAAGACCATTCGAGAACATTTAACGAGCTATGGAAGAAATCCAA	789
Db	1138	TACATTTAAAAAATTAAAAAAACCATAGAAATATTAATGATTTAATTTGAAGAAAGTAAAG	1197
Oy	790	AAGACCATACACAAAATTAAGATTCGACACCAAGAGGAAGAAAGAGAGAGTGTACCAAG	849
Db	1198	AAAACAATTTATTAATAAATAAGATGCACTTAAGAAAGAAAGAAAAAAATTTATACCAA	1257
Oy	850	GCCAGTAGACACTGTCCATCTTTAACAAACAGCTTGAAGAGCCCTATACCTCATACAGC	909
Db	1258	GCTCAATATGATCTTCTCTATTTTCAATTAACATTAAGAAAGACATTAATTTAATAAGC	1317
Oy	910	GTACTGAGAGCGCATAGACACCTCAAGAAAGATGAAATATCAAGAACTGCTGCAC	969
Db	1318	GTTTAAAGAAAACGATTTGACACTTTAAAAAAATTAAGAAACATTAAAGCAATTTACTGAT	1377
Oy	970	AAGATTTAATGAATTTAAGATTCCTCGCGACGCACTCGGGAACACCCCTACACAGCTG	1029
Db	1378	AAGTTAATTAATGAATTTAATAAATCCCGACCGGCAATTTGGAATAATACCAAATATCTCTC	1437
Oy	1030	CTGACACAGAACAAGATAGAGAGACAGCAAGCAAGAAAGATCAAGAGATGCCAAAAC	1089
Db	1438	CTTGATTAAGAACAAAAATCGAGGAACACGAAAAAGAAATTAAGAAATTTGCCAAAAC	1497
Oy	1090	ATTAAATTCACATAGATTTCTCTTTACTGATCCCTGTGAGCTGGAGTACTTACTTGAGA	1149
Db	1498	ATTAAATTTATATTAATGATTTTATTTACTGATCCACTGAATTTGAATTAATTTAAGA	1557
Oy	1150	GAGAAATTAAGATTTATAGACATCTCCGCAAGTCGAGCAACAAAGATCAACGAACT	1209
Db	1558	GAAAAAATTAATTAATTTGATTAATAGTGCMAAGGTTGAACAAAGAAATCAACTGAAACC	1617
Oy	1210	AATGAATATCCCAATGCTGTGACGTAACCCCTTCGATTTACGAATATCAACAGCTCTC	1269
Db	1618	AATGAATATCCCAATGAGAGTACTATCTCTTGTCAATTAACGATTAATCAATAGCTTTA	1677
Oy	1270	AAGAGCTCAATGCTGTGCTGCTGATTAACCCCTTCGATTTACGAAGAAACCTCT	1329
Db	1678	AATGAATTAATCTTTTGTGTGATTTAATTAATCCATTTGATTAACAAAAGAACCAAGT	1737
Oy	1330	AAGAAATTCACACAGACATAGCAAGAAAGAGTTTATCAACGAATCAAGAGAGATC	1389
Db	1738	AAAAACATATATCTGATATAGAAAGAAAAAATTCATTAATGAATTAAGAAAAAATTT	1797
Oy	1390	AAATTTAGAGAAAGAAATTTGAGAGTGCACAAAGAAAGTTACGAAGACCCGACGAAAAAGT	1449
Db	1798	AAATTTAGAAAAAAGAAATTAATGATTTGATTTAAAAAATCTTTACGAAGACAGATTTCTAAGTCT	1857

QY	1450	CTAACGATATCACTAAAGAGTATGAAAAGCTCTCTAACAGAGTCTATGATTCCAAATTC	1509
Db	1858	TTTAAATGATATACAAAAAGAAATGAAAAATTCCTTAAGAAATTTATGATGCAAAATTC	1917
QY	1510	AACATTAACATCGAGCTGACCACCTTGACAAAATGATGGGAAAAACGGACTCTACANA	1569
Db	1918	AATATTAATATAGATTTTAATTAATTTCTGAAAAAATATGGTAAAGATATTCTATATAA	1977
QY	1570	GTGGAAAACACTGACACACCATTAATACCTTTCATCTATAGAAATTTCTAAGCATATCTT	1629
Db	1978	GTTGAGAACTTACACACCATTAATACCTTTGCATCTTATGAAAAATTTCTAATCATATCTT	2037
QY	1630	GAGAAAGCTCACCAAAGCTCTTAAAGTATATGAGAGCATTTCTCTCGGAACATTTTGTG	1689
Db	2038	GAAAAATTAAACAAAAGCTTTAAATATATGGAAGATTTATTTCTTAAAGCATATACTACTT	2097
QY	1690	GAGAAAGAACTAAAGTATTTCAAGAATCTCATAGTAAATCATGGAAGACAGATTCGACAGC	1749
Db	2098	GAAAAAGATTTAAATTTTATTAATAATTTATTAAGCAAAATGAAAATGAGATTGAAACA	2157
QY	1750	CTTTGTGGAACATTTAAGAGATGGAAGAACACTGTTTGTGAGAGAGATTTACAAAAGAC	1809
Db	2158	TTAGTTGAAATATTTAAAAAAGATGAAGAACACTTTTGTAAAAAAAATTAATCTAAAGAC	2217
QY	1810	GAAATTAACACCATGATGGAAGATCTCGGAGCTCCGATATTGTTTAAAGTCCAAAGTCAG	1869
Db	2218	GAAATTAACACCATGATGGAAGATTTTAAAGATATCTGACATTTGTAAAGTCAAACTCCA	2277
QY	1870	AAGGTCCTCATGAAACAAAGATTGATGAACTCAACACATCACTCATCTGGAAGAC	1929
Db	2278	AAAGTTTATTAAGAACAAATTTGACGAATTTAAAAAAGATCAATGATTTTAAAAAAT	2337
QY	1930	GTGAGATTAACAACTAATATATACATGTGCCGAATAGTTATTAACAGAGAAATTAACAGGAA	1989
Db	2338	GTAAGATTAACAACTAATATATACATGTGCCAATTTCTTCAAAACAAAGAAATTAAGCAAGAA	2397
QY	1990	CCATATCTACTATCTGATCTGATCTCAAGAAAAGAGATAGCAAACTGAAAGTTCATGCCAAA	2049
Db	2398	CCTTATTTATTAATTTGTTGAAAAAAAAGAAATTTGATTAATTTAAAGTGTCTAGCTCAAG	2457
QY	2050	GTCCAGAGCCGTATCAACAGCAAGAGAAAGAAATTAATTAACAGAGAGACGTCAAGATAC	2109
Db	2458	GTAAGATCAATTAAGATTAATTAAGAAAAAAAACAAATTAACAGAAAGTCAATCGGATATAT	2517
QY	2110	TCCGAGCCCTTCACAGAGAAGAGATTAACCGGACAGCTTACCCACCAACCCGAGCAACAG	2169
Db	2518	TGCGAACCATTAACCCGAAAGGAAATTAACAGCAACAGCAACTTACAAAACCTGTGACACAA	2577
QY	2170	GCCGGTTACGCTCTCGAAGCGATAGCGTGCACAGCTCAAGCAACAGACAGAGCAAGCGCA	2229
Db	2578	GCAGGATCTGCTTTAGAAAGAGATTCGATCAAGCAACAGCAAGCAAGAAACAAAACAGCA	2637
QY	2230	CAGCTTCAGTGGCGAGTGGCCGTTCCAGAGGCTTAAAGCTTCAAGTGGCTTACACACACAGCT	2289
Db	2638	CAACCCACAGTACACAGTACCAAGTACCAAGAGCAAGGCAAAAGCAAGTCCCAACACACACACA	2697
QY	2290	CGTGTGAATTAACAGACCGAGATGTGCAGAAAGCTGGACTTACCTGAGAGAGCTGTATGAG	2349
Db	2698	CCAGTAAATTAATTAACCTGAATTAATGTTTCCAAATTAAGATTAATCTTGAAAAAATTAATATGA	2757
QY	2350	TTTCTGGAATACATCTCAATCTGCAACCAATATATCTCTGCTCTCACAGCAGTATGAC	2409
Db	2758	TTTTTAAATACCTTATATATATATGTCACAAATATATTTTGGTTTCAACCTCAAGTATGAC	2817
QY	2410	GAGAAGATTTTAAACAGTCAAGATTAACCAAGGAAGAGAGATTAACCTGTCTTGT	2469
Db	2818	GAAAGATATTAACAACTAATTAATAATTAACAAAGAGAGAAAGCAAAATTAAGTTCATGCT	2877
QY	2470	GATCACAGGACCGCGTTCATATTCGAAACACACATTCCTCCGTATGATATTTATGATC	2529
Db	2878	GATCATATTAAGCTTATTTTAAATTAACAAATTAACATACCTGTATGATTTATTTATGTTT	2937
QY	2530	GATAGCCTCAACATTTCTCTCTCAACGTTTCATGAGATTTATGAACAGAGATGGTC	2589

DB 2938 GATAGCTTAACAATAGTTATTCACAACATTTATGAAATTTATGAAAAAGAAATGTT 2997
QY 2590 TCGAACCTGATATACCAAGACAGACAAAGATTAAAGACCTTCGGAGGAAGCTAAG 2649
DB 2998 TGTAAATTTATTAACCTTAAGATTAATGACAAAAATTAATTAATTAAGGAGAAAGCCAAA 3057
QY 2650 AAGGTCTCCACTCTGTAAACTCTCTCCAGCTCCATGCAACCAAGCTCTCTGACA 2709
DB 3058 AAGGTATCCACATCTGAAAAACTCTTCAAGTTTCATCATGCAACCATTTATTCATTACA 3117
QY 2710 CCTCAGACAGACCCGAGTAGAGGCTAACGAGACACCTCTCACTGACCAACTTAAT 2769
DB 3118 CCTCAGAGATTAACCCGAGTAGAGGCTAACGAGATGATCATCATCTCTCAAAATTTGAAT 3177
QY 2770 AACTCACTGAAAGCTGTGGAACATCCCTCTGCGGCAAGATATAGAATCTTACCAA 2829
DB 3178 AATAGATTTAAATATTTATTAAGAAACATTTAGATCTTGAAAAAACAATAATATATCCAA 3237
QY 2830 GAACCTATTGGAGAGAAATCGTCGAGAACCTTCTACGAGAGATACTGAAGACAGCGAC 2889
DB 3238 GAATTAATAGTCAAAAAGATAGAAAACCTTTATGAAAAAGATTTAAAGATAGTGAT 3297
QY 2890 ACATTTCTATACGAGAGCTTCACTACTTGTGAAATCTTAAGCCCATGATATCACTCT 2949
DB 3298 ACATTTTATATGAAATCTTTTACAAATTTTGTAAATCTTAAGCTGATGATTAATTTCA 3357
QY 2950 CTTAACGATGATCTTAACCTAAGAACCTGAGAGACATCAATAGCTGAAGAGACA 3009
DB 3358 TTGAATGATATATCAAAAAGAGAAATTTAGAAAGATTTAATTAATTAATAAAAAAACT 3417
QY 3010 CTGCAACTGAGCTTCGACCTGTACAAACAGTACAACTGAGAGACTCTTCGAC 3069
DB 3418 TTACAGATTAATCATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTTGAT 3477
QY 3070 AAGAAGAGACACTCGGCAAGTAAAGATGCAAGATCAAGAAAGTTGACTGCTCAAGAG 3129
DB 3478 AAAAGAGAAACAGTGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAAAGAA 3537
QY 3130 CAGCTTGAAGACAACTCAACTGCAAGCAATCCGAAACAGCTACTGCAAGACTTCTCA 3189
DB 3538 CAATTAAGATCAAAATTTGATTCATTAATTAACCAAGATGATTAACAAACTTTCT 3597
QY 3190 GTGTTCTTCAAGAAAGAAAGAGCCGAGATGCGGACAGAGAACTCTGAGAGAC 3249
DB 3598 GTTTTCTTCAAGAAAGAAAGAGAGGAAATAGCAAAATGAAACATTAATTAAGAAAC 3657
QY 3250 ACCAGATTTCTTCAAGACATCAAAAGGCTGTCAGATTAATTAATTAATTAATTAAT 3309
DB 3658 ACAAAATATTTATTTGAAACATTTAATTAAGACTGTTAATTAATTAATTAATTAATTAAT 3717
QY 3310 CCTCTGAAGACTCTCCGAGAGAGACATCCAGACGAGATTAATTAATTAATTAATTAAT 3369
DB 3718 CCATTAATAAATTTAAGTGAAGATCAATTCAAACAGAAATTAATTAATTAATTAATTAAT 3777
QY 3370 AACTTCAGAGTCTGTCTAAGCTGCAAGGCAAGCTGAGAGCAACTGAACTGAGAGAG 3429
DB 3778 AACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3837
QY 3430 AAGAAGCTGATACCTCTCTAGCGAGCTGATCACTGATCCCGAGCTCAAGAGAGTC 3489
DB 3838 AAAAATTTATCATCTATTCATCAAGTGAATCAATTAATTAATTAATTAATTAATTAAT 3897
QY 3490 ATTAAGAAACAAGACTACAGCCGATAGCCCAAGGAGAGATTAATTAATTAATTAATTAAT 3549
DB 3898 ATTAATAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3957
QY 3550 GCACCTGGAATCTTACAGAAAGTCTGCTGAGAGAAACAGATGCTGCGCACTGCTGCTCT 3609
DB 3958 GCATTTGAATCTTACAAAAATTTCTCCGAGAGAAACAGATGTTGCAACAGTTGTAGT 4017
QY 3610 GAATCTGGCTCCGACACTGAGAGAGTCTCAACCTTAAGAGCCTGATCTACTCATGTC 3669

DB 4018 GAAAGTGATCCGACACATTAAGAACAAAGTCAACCAAGAAACAGCATCACTCATGTA 4077
QY 3670 GGAGCGAGTCCAAATCAATTAACCATCTCAGAGCTGAGAGTGAAGTGCATGTC 3729
DB 4078 GGAGCGAGTCCAAATCAACAAATTAACCAATCAACAAATTCGATGTAAGTATGACGTA 4137
QY 3730 ATCATGTGCTCTATCTTCCGAG 3789
DB 4138 ATCATAGTACTATTAATTTGAGAGATCCGAGAGAGATTAATGATGATTAAGACAAGTAGTA 4197
QY 3790 ACCGCTGAGCTGCTACTCTTCCGATGATGATTAATCTTGTCCAAATTCGAGAGAGAG 3849
DB 4198 ACAGAGAAACAGTAACTCTCTCCGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4257
QY 3850 TACGAAGTCTATCTATGAAACCTCTGAGAGGCTCATGATGCTCTCAAGAAACAGCTG 3909
DB 4258 TATGAGTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4317
QY 3910 GAGATTAAGCTGATGACCTTCAATGTCACAGTGAAGAGATTTGAAACAGCCCTTTAAT 3969
DB 4318 GAAATTAAGCTATGACATTTAATGTTAATGTTAATGATTTAATTAATTAATTAATTAATTAAT 4377
QY 3970 AAGAGAAATTTCAAGAAAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4029
DB 4378 AAAGTGAATTTTCAAAAATTTTGAATTCAGATTTAATTAATTAATTAATTAATTAATTAAT 4437
QY 4030 TCCCTCACTAGCTTGTCAAGAGCCATACAGTCTCTCATTAAGAGAGAGATTAAT 4089
DB 4438 TCAAGTATTAATGTTGTCAAGAGATCCATTAATTAATTAATTAATTAATTAATTAATTAAT 4197
QY 4090 TTTCTGTCTAATCAATTAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4149
DB 4498 TTTCTTAAGCAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4557
QY 4150 GATGCTGCTGGATTAATCAAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4209
DB 4558 GATGCTGCTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4617
QY 4210 AAAAGATTAATCAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4269
DB 4618 AAAAGATTAATCAAG 4677
QY 4270 ATGGAAGCTCTGTAAG 4329
DB 4678 ATGGAAGCTCTGTAAG 4737
QY 4330 GCCAAGCTCTCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4389
DB 4738 GCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4797
QY 4390 AACTACCTCAAAACATCCAG 4449
DB 4798 AATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4857
QY 4450 GGAATTCAGAGCTGTCTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4809
DB 4858 GGAATTCAGAGCTGTCTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4917
QY 4510 ACTGCGATGCTGTTGAAACCTCTGCAAAACAGAGTCTGAGCATCTGCTGCAAGAGAGTC 4569
DB 4918 ACAGGATGCTGTTGAAACCTCTGCAAAACAGAGTCTGAGCATCTGCTGCAAGAGAGTC 4977
QY 4570 CTGCAAGGCTGCTGTAACATCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4629
DB 4978 TTGCAAGGATGTTAAACATTTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5037
QY 4630 AGCGGCTGTTTCAAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4689
DB 5038 TCTGAGATGTTTCAAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5097
QY 4690 CAAGAGAGAGATTAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4749
DB 5098 CAAGAGAGAGATTAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5157

JOURNAL Patent: WO 9814583-A 1 09-APR-1998;
PAN WEIQING (DE); BUJARD HERRMANN (DE)
FEATURES Location/Qualifiers
source 1. 4920
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2194 a 589 c 679 g 1458 t
ORIGIN
Query Match 57.0%; Score 2814.4; DB 6; Length 4920;
Best Local Similarity 73.3%; Pred. No. 0;
Matches 3604; Conservative 0; Mismatches 1316; Indels 0; Gaps 0;
QY 10 ATGAAATCATTTCTCTCTGTCATTTCTGTTTTATCATCAATCTAGTGGTG 69
Db 1 ATGAGATCATATTTCTTTTATGTTCAATTTCTTTTATTAATAATACAAATGTGTA 60
QY 70 ACCCGCAATCTATCGAGAGCTGTGAAGAACTGGAAGTGGAGTGGCTCTT 129
Db 61 ACACATGAAAGTTATCAAGAACTTGCAAAAACCTAGAACTTGAAGATGCGATTTG 120
QY 130 ACCGATACAGCTGTCCAGAAAGAGATGGTCTGATGGAAGAGAGAGTGGACG 189
Db 121 ACAGGTTATGTTTATTTCAAAAAGAAAATGTTATTAATGAAGAAACAAGTGGACA 180
QY 190 GCGGTTACAACACGACACCCGGTTCTAAAGGCTGTGCTAGCGGTGGCTCGGTGG 249
Db 181 GCTGTACACATAGTACACCTGGTTAAAGGGTTCAGTTCTTCAGGTGGTTCAGGTGGC 240
QY 250 TCTGTGGCTCTGGGGTTCCTGCTCCCTCCGGGGGCGACGCTGGCATAGGTGGTCA 309
Db 241 TCAGTTGCTCAGGTGGCTCAGTTGCTTCAGGTGGCTCAGGTGGCTCAGTT 300
QY 310 GCAAGGCGGCTTCGCGGAACAGTCGAAAGAACCAATCTGACACTAGCATTC 369
Db 301 GCTTCAGGTGGTTCAGTATTCAGAGCTACCAATCTTCAGATTAATTAAGGATTTCA 360
QY 370 GACGCCAAGTCTGACGCCGACTCAAGCACCAGGTGAAGAACTATCTCTCACTATCA 429
Db 361 GATGCTAAATCTTACGCTGATTTAAACACAGATGACGAATTAATTAACATCAAA 420
QY 430 GACGTGAAGTACCCAGCTGTGCACTGACCTCAATCATATGCTGACACGCTGATAC 489
Db 421 GAACCTAAATATCTCACTCTTGAATTAACATCATATCTTGAATTTGATGATAT 480
QY 490 ATTCATGCTCAATATCTGATTTGAGGCTTACGAGATGATCAATGAACTCTGTACAG 549
Db 481 ATTCATGCTTCAATATTTAATTTAGATGATGATGAAATTAATTAATTAATTAATAA 540
QY 550 TTGAATTTCTACTTCGACTGTAGAGGCCAACTGATGACGTTGCGCAATGACTAT 609
Db 541 TTAACTTTTATTTGATTTTAAAGCAAAATTAATGATGATGCTTAATGATTAAT 600
QY 610 TGTCAATTCATTCATTAATTTGAGATGAGCAACGAGTTGACGATGGAAGAAATTG 669
Db 601 TGTCAATTCCTTCAATCTTAATAATTCGCAAAATGATTAAGCGTAACTTAATAA 660
QY 670 GTCTGCGATATGCAAGCCTCTGCAACATCAAGGACATGTGGAAGATGGAAGAT 729
Db 661 GTCTGCGATATGGAAGAACCATTTAGACATTTAAGATATATGAGAAAAATGGAAGAT 720
QY 730 TATATTTAAAAAGATTAAGAAACCATCGAGACATTTAAGAGCTGATGAGAAATCCAAA 789
Db 721 TATATTTAAAAATTAAGAAACCATGAGAAATTAATTAATTTAAGAAAGTAAG 780
QY 790 AAACCATTAACAATAATAGATGCAACGAGAGAGAAAAAGAAAGTGTACAG 849
Db 781 AAACCATTTGATTAAGAAATGAATGCACTTAAGAAAGAAAAAATTAATTAACAA 840
QY 850 GCCCAGTACAGCTGTCATCTATTAACAAACAGCTTGAAGAGGCCCTAATCCTATCAGC 909
Db 841 GCTCATATGATCTTCTTATTTACATTAACCAATTTAGAAAGACACTTAATTTAATTAAGC 900

QY 910 GTACTGAGAAAGCCATATAGACACCCCTCAAGAGATTAATAATATCAAGACACTGCTGAC 969
Db 901 GTTTATAAAAACGCTATGTGACACTTTAAAAAATTAATAAGATTAATCTGAT 960
QY 970 AAGTATATGAATTAAGAAATCTCCGCGACCACTCTGGGAACACCCCTAACAGCTG 1029
Db 961 AAGTATATGAATTAATAATCCCGCACGCGCAATCTGTGAAATTAACCAATATCTCTC 1020
QY 1030 CTGACACAAGAAACAAGATATAGAGAGACAGAGAAAGATCAAGAGATCCCAAAAC 1089
Db 1021 CTGATATAGAAACAATAATCGAGAGACAGAAAAAGAAATTAAGAAATTTGCCAAACT 1080
QY 1090 ATTAAGTTCAACATAGATCTCTCTTACTGATCCCTTGAGCTGGAGTACTCTGAGA 1149
Db 1081 ATTAATTTATTAATTTGATTTATTTATTTATTCAGATCCCATTAATTAAGATTAATTTA 1140
QY 1150 GAGAGATTAAGATATAGACATCTCCGCAAGTGGAGCAAAAGATCAACGCAACT 1209
Db 1141 GAAAAAATTAATAATATGATATTAAGTGCAAAAGTTGAACAAAGAAATCACTGAACCC 1200
QY 1210 AATGAATATCCCAATGCTGACGTACCTCTGCTTATACATATCAACAAAGCTCTC 1269
Db 1201 AATGAATATCCCAATGAGATGATCTTATCTTGTATATACGATATTAACAAATGCTTTA 1260
QY 1270 AACGACTCAATAGCTTCGCTGACTGATTAACCCCTTCGATTTACGAAAGAACCTCT 1329
Db 1261 AATGAATCTTAATCTTTGCTGATTTAATTAATTCATTTATTAATTAACAAAGCAAGT 1320
QY 1330 AAGATATCTACACAGCAATGAGAGAAAGATTTTCAACGAAATCAAGAGAAATC 1389
Db 1321 AAAAAATATATCTGATTAATGAAGAAAAAATTCATAATGAATTAAGAAAAATTT 1380
QY 1390 AAAATGAGAAAGAAAGAAATTTGAGAGTACAGAAAAAGTTACGAAGACCCAGCAAGAT 1449
Db 1381 AAAATAGAAAAAATAATTTGATCTGATTAATAAATCTTACGAAGACAGATCAATGATC 1440
QY 1450 CTAAAGATATCTTAAGAGTATGAAGAACTGCTGACAGATCTATGATTTCCAAATTC 1509
Db 1441 TTAATATATTAACAAAGATATGAAGAAATTTCTTATTAATAATTTGATGATGAAATTC 1500
QY 1510 AACATATCATCTGACCTGACCACTTGAGAAAAATGATGGAAAAAGCTGCTTACAA 1569
Db 1501 AATATATATTTAGTTTAACTAATTTGAAAAAATGATGGTAAAGATTTTCAATATAA 1560
QY 1570 GTGAGAAACTGACACACCACTAATACCTTTGCATCTATGAGAAATTTCAACATATCT 1629
Db 1561 GTTGAGAAACTTACACACGATTAATACCTTTGCATCTATGAAATTTCAACATATCT 1620
QY 1630 GAGAGCTCACCAAGCTTTAAGTATATGAGAGACTATTTCTGCGGAACTTTGTTGTG 1689
Db 1621 GAAAAAGTTAAACAAAGCTTTAATAATATATGAGATTAATTTCTTTAAGCAATTAATGAT 1680
QY 1690 GAGAAAGCACTAAAGTTTCAAGAAATCTAATGATTAAGTGAAGAAAGGATGCAAGC 1749
Db 1681 GAAAAAGATTAATAATTTTAAAAAATTTAATAGCAAAATAGAAATGAGATTTGAACA 1740
QY 1750 CTTGTTGAGAACATTTAAGAAAGTGAAGAACAGTTGTTTGAAGAAAGATTTAACAAGAC 1809
Db 1741 TTAGTTGAAAAATTTAAAAAAGATGAGAAACAGCTTTTGAAAAAAATTTACTAAGAC 1800
QY 1810 GAAAAATTAACCAAGATGAGAGATCTGAGAGCTCCGATATTTGTTAAATGCCAAGTCCAG 1869
Db 1801 GAAAAATTAACCAAGATGAAAAATTTTGAAGTATCTGCAATTTGTAAGAAAGTCAATTC 1860
QY 1870 AAGGTGCTCTCATGAACAAGATTTGAGAGCTCAAGAGAGCTCAACGATTTGGAAGAC 1929
Db 1861 AAGTTTATTAATGAACAATAATGAGCAATTTAAGAAAGCTCAATGATTTTAAAAAAT 1920
QY 1930 GTGAGTTTAAACATATATATACATGTGCGCAATAGTATTAAGAGAGAAATTAAGCAAGAA 1989
Db 1921 GTGAATTTAAACATATATATACATGTTCCTCAATTTTACAAACAGAAATTAAGCAAGAA 1980
QY 1990 CCATTAACCTCATCTTACAGAAAGAGATAGACAACTGAAGATGTTCAATGCCCAA 2049


```
Db 1981 CCTATTATTTAAATGTGTGAAAAAAGAAATATGATTAATTTAAAGTGTTCATGGCTAAG 2040
Qy 2050 GTCGAGACCTGATCAGCAAGAGAGAAAGACATTTAAACTGAAGCAGCTCAGATAAC 2109
Db 2041 GTAGATCATTTGATTAATGAAGAAAAAACAATAAAAACAGAAAGTCATACGGATTAAT 2100
Qy 2110 TCCGAGCCTTCACAGAGAGAGATTAACCGAGCGTACCAAGCCGAGCCGAGCAACAG 2169
Db 2101 TCGGACCATTAACCGAGAGAGAAATTAACAGCACAACCTACAAACCTGCGACACAA 2160
Qy 2170 GCCGGTTCAGCTTCGAGGCGATAGCTGCAAGCTCAAGCAAGAGCAGAGAGAGCA 2229
Db 2161 GCAGATCTGCTTTAGAGGAGATTCAGTACAGCACAAGCACAAGAAACAAAAACAGCA 2220
Qy 2230 CAGCTCCAGTGCAGTCCCGCTTCAGAGGCTAAAGCTCAAGTCCCTACACCACAGT 2289
Db 2221 CAAACCACGATCCAGTCCAGTACCAAGCAAGCAAGTCCCAACACCACAGCA 2280
Qy 2290 CCTGTGAATTAACAAGACGAGATGTCAAGAACTGACCTTGAGAGGCTCTATGAG 2349
Db 2281 CCAGTAAATTAATAACGTAAGAAATTTTCCAAATTAGATTATCTTGAAAAATTAATGAA 2340
Qy 2350 TTCCGAAATACCTCTACATCTGCACAAATATATCTCTCTCAGAGCAGTATGAA 2409
Db 2341 TTTTAAATCTCTCATATATATGTCAAAATATATTTGGTTTCACTCACTCAATGAA 2400
Qy 2410 GAGAAGATTTCTTAACAGTACAGATTAACCAAGAGAGAGAGATTAAGTCTCTTGT 2469
Db 2401 GAAAAAATTTAAACAATTAATAATTAACAAGAGAGAGAAAGCAAAATTAAGTTCAGT 2460
Qy 2470 GATCCACTGAGACCTGCTGTCAATATCAAGAACATTTCCCGTATGATTTATGTTTC 2529
Db 2461 GATCCATTAGACTTTATTTTATATATACAAATTAACATACCTGTAATGATTTCTTGT 2520
Qy 2530 GATGACCTCAACATTTCTCTCTCACTGTCATGAGATATATAGAGAGAGATGTC 2589
Db 2521 GATGACCTTAACAATAGTTTATACACACTATTTATGAAAAATTTAGAAAAAATGTT 2580
Qy 2590 TGCACCGTATTAACCAAGAGACAAAGCAAAATTAAGAACCTCTGAGAGAACTAG 2649
Db 2581 TGTATTTATATTAACCTTAAGGATATATGCAAAATTAATAATTTATGAGAGAGCA 2640
Qy 2650 AAGGTCTCCACCTCTGTTAAACTCTCTCCAGCTCCAGCAACCACTGCTCTCA 2709
Db 2641 AAAGATCCACATCTGTAAACCTCTTCAAGTATCATGCAACCAATTAATCA 2700
Qy 2710 CCTCAGACAAAGCCGAGAGGCGCTAACGAGACACCTCTCACTGACCAACCTTAAT 2769
Db 2701 CCTCAGATTAACCCGAGAGTAAAGTCAAAATGATGATCATCATTTCAAAATTTGAAT 2760
Qy 2770 AACTCAGCAAGCTGTGAGAACATCTGTCTCTCGCGCAAGATTAAGAACATCTACCA 2829
Db 2761 AATAGTTTAATTAATTTGAAACCAATATGATGCTTTGAAAAACAAAAATTAATACAA 2820
Qy 2830 GAACATTATGAGAGAAATCTGCGAGAACTTCTACGAGAGATCTGGAAGAGCGGAC 2889
Db 2821 GAATTAATAGTCAAAAAAGTAGTAAACCTTTATGAAAAATTTTAAAGATAGTGT 2880
Qy 2890 ACATTCTTAACGAGAGGCTTCACTAACTTCTGAAATCTAAAGCCGATGATTTCACT 2949
Db 2881 ACATTTTATATGATCTTTTACAAATTTTGTAAATCTAAAGCTGATGATTTATTTCA 2940
Qy 2950 CTTAACGATGATTAACGTAAGAGCTGAGAGAGACATCAATTAAGCTGAAGAGCA 3009
Db 2941 TTGATGATGATCAAAAAAGAGAAATTAAGAGAGATTAATTAATTAATAAACT 3000
Qy 3010 CTGACAGTACGCTGAGCTGTAACAACAAGTAAAGAACTGAGAGACTCTTCAC 3069
Db 3001 TTACGTTATATTTGATTTTATATATTAATTAATTAATTAATTAATTAATTTGAT 3060
Qy 3070 AAGAAGAGACGTCGCAAGTATTAAGATGAGATCAAGAGTTGACTCTGCTCAAGAG 3129
Db 3061 AAAAAAGAACGTTGTAATATATAATGCAAAATTAATAAACTTACTTATTAAGAA 3120
Qy 3130 CAGCTTGAAGCAAACTCACTGAAACAATCCGAAACGCTACTGAGAACTTCTCA 3189
Db 3121 CAATTAGATCAAAATGTAACTTATTAACCCAAAGCAGTATTAATAAACTTTCT 3180
Qy 3190 GTGTTCTTCACAGAGAGAGAGCCGAGTCCGAGACAGAGAACCTCTGGAGAAC 3249
Db 3181 GTTTTCTTTACAAAAAAGAGAGAGAGTAAATAGCAGAACTGAAACACATTAAGAAAC 3240
Qy 3250 ACCAAGATTTCTTCAAACTACAAAGGCTGTCAGATTAATTAATGCGAGTCTCT 3309
Db 3241 ACAAATATTAATTAAGTAACTTATTAAGAGCTTTGTAATTAATTAATTAATTAAT 3300
Qy 3310 CCTGGAAGACTCTCCGAGAGAGATCCAGACCAGAGATTAATTAATTAATTAATTAAT 3369
Db 3301 CCATTAAAACTTTAAGTAAAGTAAATCAATTAACAGAGAGATTAATTAATTAATTA 3360
Qy 3370 AACTTCAAGTCTCTGTAAGCTCGAAGGCAAGCTGAAGCACTGAACTGAGAGAG 3429
Db 3361 AACTTAAAGTATTAAGTAAATTAAGAGAAATTAAGAGATTAATTAATTAATTAATTA 3420
Qy 3430 AAGAAGCTCAGCTACCTCTTAAGCGACTCATCACTGATCGCGAGCTCAAGAGATC 3489
Db 3421 AAAAAATTAATCACTTATTAAGTAAAGTAACTTATTAATTAATTAATTAATTAATTA 3480
Qy 3490 ATTAAGAACAACTACACCGGCAATAGCCCAAGCAGAAATTAATTAATTAATTAATTA 3549
Db 3481 ATTAATAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
Qy 3550 GCAGTGAATCTTAAGAGAGTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3609
Db 3541 GCATTAGAACTTACAAAAAATTTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Qy 3610 GAATCTGCTCGACACACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3669
Db 3601 GAAATGAGATCCGACACATTAAGCAAAAGTCAACCAAGAGAGAGAGAGAGAGAGAGAG 3660
Qy 3670 GGAGCGAGTCCATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCA 3729
Db 3661 GGAGCAGAGTCTTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCA 3720
Qy 3730 ATCATTTGCTATCTTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3789
Db 3721 ATCATTAAGTACCTAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
Qy 3790 ACCGCTGAGCTGCTACCTCTTCCGATTAATTAATTAATTAATTAATTAATTAATTAAT 3849
Db 3781 ACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Qy 3850 TACGAGAGCTATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3909
Db 3841 TATGAGGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3900
Qy 3910 GAGAAATACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3969
Db 3901 GAAATTAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3960
Qy 3970 AAGAGAGAAATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4029
Db 3961 AAGCTGAGAAATTTCAAAATTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Qy 4030 TCCCTTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4089
Db 4021 TCAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4080
Qy 4090 TTTCTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4149
Db 4081 TTTCTTAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4140
Qy 4150 GATGCTGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4209
Db 4141 GATGCTTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4200
```


Db	636	ATTCTGGCTTCAAAATATTTTAATTTGATGGATATGAGAAATTTAAGAATTTATTAATATAA	755
Qy	550	TTGATTTTCTACTTCGCATCTGGCTAAGGGCCAAACTGAATGAGCTTTGGCCCAATGACTAT	609
Db	756	TTAAACCTTTATTTTGTATTTATTAAGAGCAAAATTAATAGATGTATGTCTATATATATAT	815
Qy	610	TGTCAAATTCATTCCAAATTTTGAAAGATCAGACCAACGAGTGTGACGTATTTGAAGAATTG	669
Db	816	TGTCAAATTCCTTTCAATCTTTAAATTCGTGCAAAATGAATTTAGAGCTACTTTAAAAACCTT	875
Qy	670	GCTCTGGATATGCGCAAGCCCTCGCAACATCAAGACAAATGTGGAAAGTGTGAACAT	729
Db	876	GTTTCGGATATACAAAACCATTTAGACAAATTTTAAAGATTAATGTAGAAAAATTTGAAACAT	935
Qy	730	TATATTTAAAAAGAAATTAAGAACCATTCGAAACATTATACGACCTGATCGAAGAAATCCAA	789
Db	936	TACATTTAAAAAAATTTAAAAAACCATGAAATATATAATGAATTTATTTGAACAAAGTAA	995
Qy	790	AAGACCATAGACAAAATTAAGATGCAACCAAGAGAGAAAGAAAAGAAAGTTTGACAG	849
Db	996	AAAACAATTTGATTAATAAATAAGAAATGCACTAAAGAAAGAAAAAAATTTATACAA	1055
Qy	850	GCCCACTGCGACGCTGCCATCTAATACAAACACTTGAAGAAAGCCCATTAACCTCATCAC	909
Db	1056	GCTCATATGATCTTCTCTATTTCATTAACAAATTAAGAAAGAACATTAATTTAATAAGC	1115
Qy	910	GTACTGGAAGCGCATAGACACCCCTCAAGAAATGAAGAAATATCAAGAACTGCTGCAC	969
Db	1116	GTTTGAAGAAACGATATTGACACTTTAAAAAAATTTGAAAACATTTAAGAAATTAATCTAT	1175
Qy	970	AAGATTAATGAATTTAAGAAATCCTCGCGCACGCACTGTGGAGAACCCCTTAACACGCTG	1029
Db	1176	AAGATTAATGAATTTAAAAATTTCCGCCACGCGCAATTTGTGAATTAACAAATTAATCTCTC	1235
Qy	1030	CTGACCAAGAACAGAAGATATGAGAGACACGAAAGAATCAAGAGATCGCCAAACC	1089
Db	1236	CTTGATTAAGAACAAAAAATCGAGGAACAGAAAAGAAATTAAGAAATTTCCCAAACT	1295
Qy	1090	ATTAAGTTCAACATAGATTTCTCTTTTACTGATCCCTTGACCTGGAGTACATCTTGAGA	1149
Db	1296	ATTAATTTTAATTTGATTAATTTATTTACTGATCCACTTGAATTTAGATTAATTAATTAAGA	1355
Qy	1150	GAGAAAGAAATAAGATATATGACATCTCCGCCAAAGTCGAGACAAAGAAATCAACCGAACCT	1209
Db	1356	GAAAAAATTAATAATTTGATATATAGCGAAAGTTGAAACAAAGAAATCAACGTACAC	1415
Qy	1210	AATGATATTTCCCAATTTGCTGTGACGTACCTCTGCTTATTAACGATATCAACACGCTCTC	1269
Db	1416	AATGATATTTCCAAATGGAAGTACTTATCCCTTGTCATATAAGATATTAACATGCTTTA	1475
Qy	1270	AACGAGCATACCTACCTCGGTGACCTGATTAACCCCTTGATATTAACGAAGAACCCCT	1329
Db	1476	AATGAACTTAATCTTTTGTGGATTTATTAATTAATTCATTTGATATTAACAAACCAACGAT	1535
Qy	1330	AAGAATATCTACACGACATGAGAGAAAGATTTATCAAGAAATCAAGAGAGAGAT	1389
Db	1536	AAAAACATATATCTGATATATGAAGAAAAAATTCATATAAGAAATTTAAGAAAAAATTT	1595
Qy	1390	AAAAATTGAGAGAAATTTGAGAGTGCACAGAAAGTTACGAAGACCGCAGCAAAAGT	1449
Db	1596	AAAAATGAAAAAATAAATTTGAATCGATATAAATAATCTTACGAAGACGATCTAAGCT	1655
Qy	1450	CTAAACGATATCTATAAGATATGAAGAAAGCGCTGAAGACATCTATGATCTCCAAATTC	1509
Db	1656	TTAAATGATATACAAAAGAAATATGAAAAATTTCTTAATGAAGAAATTTATATGCAAAATTC	1715
Qy	1510	AACATATAACATCGACTGACCACTCTCGAAGAAATGATGGAAAAACGGTACTCTACAA	1569
Db	1716	AATATATATATATGATTTTAATGATTTTGTGAAAAAATATGTGGTAAAGAAATTTCAATATAA	1775
Qy	1570	GTGGAAGAACTGACACACCAATTAATACCTTGTGACTTATGAAGAAATTTAAGCATATATCTT	1629

Dd	1776	GTGGAAGAACTTACACACCCTTAATACTTTGGCATCTCTATGAAAATTTCTTAACATATACTT	1835
Qy	1630	GAGAAAGCCACCAAAAGCTCTTAAGTATATGAGAGACTATTTCTCGGGAACATTTGTGTG	1689
Dd	1836	GAAAGATTAAACAAAAGCTTAATATATATGGAAGATTAATTTCTTAAGGAATPATAGT	1895
Qy	1650	GAGAAAGACTTAAGCTTTACAAAGATCTCATTAAGTAAGTGGAAAGATCGAATCGAGAC	1749
Dd	1896	GAAAAAGAAATTTAAATTTTATTAATAATTTATTAAGCAAAATAGAAATAGAGATTGAAA	1955
Qy	1750	CTGTGTGGAATCTTAAGAAGATGGAAGAACCTTTGTTGAGAGAAAGATTTCAAAAC	1809
Dd	1956	TTAGTTGAAATATTAAAAAAGATGAGAGACACCTTTTGAAAAAAAATTTCTTAAGAC	2015
Qy	1810	GAAATAAACCAGATGAGAGATCTGGAGAGTCTCCGATATTGTTAAAGTCCAAAGTCAG	1869
Dd	2016	GAAATATAACCAAGATGAAAAAATTTTGAAGATATCTGACATTTGTTAAAGTACAACTTCAA	2075
Qy	1870	AAAGTGCTCCTCATGAACAAAGATTGATGAATCAGAGACTCACTCTTCTGAAGAC	1929
Dd	2076	AAAGTTTATTAATGAACAAAATTTGACGATTTAAAAAACCTCAATGCTTTTAAAAAAT	2139
Qy	1930	GTGGAGTTAAACATTAATATACATGTGGCGAATAGTTATACAGAGAGATTAAGCAGGAA	1989
Dd	2136	GTAGATTAAACATTAATATACATGTGCCAATTTCTTAACAAACGAATAATAGCAGAGA	2195
Qy	1990	CCATACTACCATCTGATCTCAAGAAAGATAGCAAAACTGAAAGTGTTCATGGCCAAA	2049
Dd	2196	CCTTATTTATTTAATTTGGTGTGAAAAAAGAAATTTGATTAATTTAAAGTGTTCATGGCTAAG	2255
Qy	2050	GTGAGAGCCTGATCTCAACGAGAGAGAGAGAACATTTAAACTGAAGGACAGTCAGATTAAC	2109
Dd	2256	GTAGAATCATTTGATTAATGAGAAAAAACAATAAAAACCAAGAGTCAATCGGATTAAT	2315
Qy	2110	TCCGAGCCTTCCAGAGAGAGAGATTAACCGGACAGCTACCAACAAGCCGGACAACAG	2169
Dd	2316	TCGGAACCATTAACCGAAGAGAAATTAACGAGCAACGACATTAACAACTGTGGAACAA	2375
Qy	2170	GCCGGTTCAGCTCTCGAAGGCGATAGCGTGCACAGCTCAAGCAACAGAGCAGAGCAGCA	2229
Dd	2376	GCAGAGTCTGCTTTAGAAGAGATTCAGTATCAAGCAACAGCAACAGAAACAAAACAACA	2435
Qy	2230	CAGCCTCCAGTGGCCAGTGCCCGTTCAGAGAGCTTAAAGCTCAAGTGCTTACACACACACT	2289
Dd	2436	CAACACACAGTACAGTACCAAGTACCAAGAAAGCAAAAGCAACAAGTCCCAACACACACACGA	2495
Qy	2290	CCTGTGAATTAACAGAACCCGAGATGTGCAGCAAACTGGACTACTTGAAGAGCTCTATAG	2349
Dd	2496	CCAGTAATAATAAACTGAAAAATGTTTCCAAATTTAGATTACTTGGAAAAATTAATATGAA	2555
Qy	2350	TTCCGGAATACATCTCAATCTGCGCAAAATTAATCTGTGTCACAGCACTATAGAC	2409
Dd	2556	TTTTTAATTAATCTCATATATATGTGCACAAATATATTTTGGTTTCCACCTCACTATAGAC	2615
Qy	2410	GAGAGAAATTTCTTAACAGTACAGATTAACCAAGAGAGAGAGATTAACCTGTCTCTTGT	2469
Dd	2616	GAAAGAATTAATAAACAAATTAATAATTAACAAAGAGAGAGAAAGCAATTAAGTTCANGT	2675
Qy	2470	GATCCACTGACCTGCTGTTCATATATTCAGAGACACATTTCCGGTATATGATATGTTTC	2529
Dd	2676	GATCCATTAAGACTTATGTTTAAATATATACAAAATTAACATCCTGTATATGATATTCATGTTT	2735
Qy	2530	GATGAGCTCAACA-----ATTCT	2547
Dd	2736	GATAGCTTTAAACATATGTTTATCCAAACTATTTATGGCTTTATGAAAAAAGAAATTTGGTTAT	2795
Qy	2548	CTCTCTCAACTGTTCATGAGATATATGAGAAAGAGATGGTGTGCACACTGTATTAACCTC	2607
Dd	2796	GTAATTTATTAATTAATTAAGAAATTTATGAAAAAAGAAATGTTTGTAAATTAATTAATAACTT	2855
Qy	2608	AAAGACAACGACAAGATTTAAGAACTTTGTGGAGGAAGCTTAAGAGGCTCCACCTCTGTT	2667
Dd	2856	AAGATATATGACAAATTAATAAATTTATATAGAGAGCCAAAAAAGTATCCACATCTGTA	2915

OY	2668	AAAACCTCTCTTCCAGCTTCATGCACACCACTGCTCTCACAACCTCAAGACACGCCGAA	2727
OY	2668	AA	2727
Db	2916	AAAACTCTTCAAGTTTCATCATATGCAACCACTATTCATTAAACCTCAGCATAAACCCGAA	2975
OY	2728	GTGAGCCCTAACGCGACGACCTCTGCATCGACCAACCTTAATTAATCTCAGTCAAACTGTTT	2787
Db	2976	GTAAGTCGAATATGTGATPACATACATTCTACAAATTTGAAATATAGTTTAAAAATTATT	3035
OY	2788	GAGAACATCCTGTCTCTCGGCAAGAAATTAAGAACATCTACCAAGAACTATTGGACAGAA	2847
Db	3036	GAAGAACATATTGAGCTCTGGAAAAAAACAAAAATATPACCAAGAAATPAATAGTCCAAAA	3095
OY	2848	TTCTGCCAGAACTCTACGGAAGATATCTGAAGAAAGACGACGACATCTCTATTAAGAGAGC	2907
Db	3096	AGTAGTAAGAACTTTTATGAAAAATTTTAAAGATTAAGATAGATCAATTTTATTAAGAAATCT	3155
OY	2908	TTCACTAACTTCGGAATATTAAGCCGATGATATCACTCTCTTAACGATGAATCTAAA	2967
Db	3156	TTTACAAATTTTGTAAAAATCTAAAGCTGATGATTAATTAATCTTGAATGATGATCAAAA	3215
OY	2968	CGTAAAGCTGGAAGAGACATCAATTAAGCTGAAGAGACACTGCACACTGAGCTTCGAC	3027
Db	3216	AGGAAAGAAATTAGACAGAGATATTATAATTAATTAATTAATTAATTAATTAATTAATTAAT	3275
OY	3028	CTGATCAACAAAGTCAACATCAAGAAACGAGAGACCTCTCGACAAAGACAGACAGTGGCC	3087
Db	3276	TTATATATATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3335
OY	3088	AAATATAGATGATGATCAAGAAATGACTCTGCTCAGAGACAGCTTGAAAGCAAACTC	3147
Db	3336	AAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3395
OY	3148	AATCACTGAACAAATCCGAAACACGTAATGAGAGACTTCTCAGTGTCTTCAACAAAG	3207
Db	3396	AATCACTATATAACCCAAAGACATGATTAACAAAACCTTTTCGTTTCTTTAAACAAAAA	3455
OY	3208	AAGAGCCCGAGATCGCCGAGACAGAGAAACACTCTGGAACACACAGATTTCTCTCAA	3267
Db	3456	AAAGAGCTGAATATGCAAGAACTGAAAACACATTGAAAACACAAAAATATTATTGAAA	3515
OY	3268	CATTAACAAAGCCCTGTCAGATATTAATTAATGCGAGCTCTTCTCTGAGACTCTCTCC	3327
Db	3516	CATTATTAAGAGACTGTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3575
OY	3328	GAGAGAGACATCCAGACCAGAGATTAATCAACCCAGCTCGAAGACTTCAAGTCTCTCT	3387
Db	3576	GAAAGATCAATTCACAAACAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3635
OY	3388	AAGTCGAGAGCACTGAGAGAGCAACCTGAACTGGAAGAAACAGCTCAGTCACTC	3447
Db	3636	AAATTAAGAGAAATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3695
OY	3448	TCATACCGAGCTCATCACTGATCGCCAGAGCTCAAGAAATCATTTAAGACAGAACTAC	3507
Db	3696	TCAAGAGGTTTACATCACTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3755
OY	3508	ACCGGCAATACCCCAACGAGATTAATTAACAGCTGAATTAAGACCTGGAATCTTACAG	3567
Db	3756	ACAGGTAATTTCTCCAAACCGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3815
OY	3568	AAGTTCCTGCTGAAGAAAGATATGTCGACATGCTGCTCTGAAATCTGCGCGACACA	3627
Db	3816	AAATTTCTCCAGAGAGACAGATGTTTGCACACGTTGTATGTAAGAAAGTGATCCGACACA	3875
OY	3628	CTGAGCAGTCTCAACCTAAGAGCCCTGCATCTACTCATGTGTGAGCGCGAGTCAATACA	3687
Db	3876	TTAGACAAAGTCAACCAAGAAACCAACGATCAACATCATGTAGAGACAGAGCTTAACACA	3935
OY	3688	ATTACACATCTCGAAGCTGACAGATGAGGTGATGACGTCAATGTTGCTATCTTC	3747
Db	3936	ATACACATATCAAAATGTGTCATGATGAAGTATGATGACGATATCATGATGATCATATTTT	3995

QY	3748	GGCGAGACCGAGGAGACTACGATGACCTCGCCAGGTGTCACCGGTGAGGGCTGTACT	3807
Db	3936	GGAGAAATCCGAAAGAAAGATTGTGATGATTAAGCAAGTAGTAACGAGGAAGCACTACT	4055
QY	3808	CCCTCCGATGTTGATTAACATCTTCGTCCAAATCGAAGCAAGATACGAAGTGCCTGATCTG	3867
Db	4056	ACTTCGGAATTTGATTAACATACCTTTCTAAATTTGAAATGAAATATGAGSGTTTATATTA	4115
QY	3868	AAACCTCTGGCAGCGCTCTTAGGCTCTCAGAAACAGCTGGAGATATAGCTGATGACC	3927
Db	4116	AAACCTTTAGCAGGTGTTTATAGAACTTTAAAAAAACAATTTGAAAATTAACGTTATGACA	4175
QY	3928	TTCAATGTCAACGTCGAGGACATTTCTGAACAGCGGCTTAATATAGAGAAAAATTTCAAG	3987
Db	4176	TTTATGTATTTGTTAAGATATTTTAAATTCACGATTTAATTAACGTAAGAAATTTGAAA	4235
QY	3988	AAACGTTTGAGAGCGACTGATTTCCCTATTAAGACCTGACCTCTACTACTGTTGTC	4047
Db	4236	AATGTTTAGATAGATGATTTAATTTCAATATTAAGATTTAACATCAAGTAATATGTTGTC	4295
QY	4048	AAGGACCATCAAGTTCCTCAATTAAGAAAGACAGGATTAATTTCTGTAGTATACAC	4107
Db	4296	AAAGTTCATTTAATTTCTTAATTAAGAAAAAGAGATTAATTTCTTAAGCAGTTATAT	4355
QY	4108	TATATCAAGGACTCCATCGACACCGATATCATTTTGGTAATGATGTGTGGGTATTAC	4167
Db	4356	TATATTAGATTCATATAGATATACGATATTAATTTTGCAAAATGATGTTCTTGATATAT	4415
QY	4168	AAGATCCGACCGAAAAATCAAGTGTGACCTTGACTCATTAATAAAGATATATCAAGAT	4227
Db	4416	AAATATATTCGAAAAATTAATTAAGATTTGATTCATTTAATAAATATATTCACAGAC	4475
QY	4228	AAGCAAGCGAATGAAAAATATCTGCCCTTCTGAATACATCGAAAACCTGTACAAG	4287
Db	4476	AAACAGGTGAAATGAGAAATACCTTCCCTTTTAACATATTTGAGACCTTATATATA	4535
QY	4288	ACATGGAACGCAAAATTCGACCTCTTGTTATTTACCTGGAGCCAGGTCTCTCAACT	4347
Db	4536	ACAGTTAATGATTAATTTGATTTATTTGTAATTCATTTAGAAACAAATCTTAATATAT	4599
QY	4348	ACTTCGAGAGACCAATGTGGAAGTTAAATCAAGACCTCACTACCTCAAAACATC	4407
Db	4596	ACATATGGAATTAACAAACGTAGAAAGTTAAATTAAGAACTTAATTACTTAACACAT	4655
QY	4408	CACAGCAAGCTGCAGATTTTCAGAAAAATTAACAATTTGCTGGGAATTCAGACTGTCT	4467
Db	4656	CAGAGCAAAATTTGGCAGATTTTAAAAAAATTAACAAATTTGTTGGAAATGCTATTTACA	4715
QY	4468	ACCGATTTAATCCACAACATCTCTGACCAAGTTTCTGTCCACATGCGCATGTGTTGCAA	4527
Db	4716	ACAGTTTATTAACCTAATTACTTATGTACAAATCTCCTTAGTACAGTATGTTTTTGAA	4775
QY	4528	AACTCGCCAAAACAGTGTGTAGCAATCTGTGCA-CGGCAACCTGACGGCATCTGAA	4586
Db	4776	AACTTGCTAAATATCCGTTTATCTAATATTTCTGATGTGAAACCTGCAAGATGTTTAA	4835
QY	4587	CATCTCCAGACCAATGCGGTGAAGAAACGTG--CCCCAGATATAGCGCGCTTTACGG	4644
Db	4836	CATTTTCACACACCAATGCGTAAAAAAAACATGATCCAAAGAGTTCGTGATGTTTCGA	4895
QY	4645	CATCTGAGACGAGCCGAAGATGTCAAGTCTCTCACTACATACAAACAAAGAGATATAG	4704
Db	4896	CATTTAGATGAAGAGAAAGATGTAAATGTTTATTAATTTACAAACAAAGAGTGTATTA	4955
QY	4705	TGCGTGGAGAAACCAAAACCTTACCTGCAATGAAAACAATGTGCGGGGTGTACGCGGATCT	4764
Db	4956	TGTGTTGAATTTCAAAATCTACTTGTAAAGCAAAATTAATGTGGATGTGATCAAGTGC	5015
QY	4765	AAATGACACCGAGAAAGACAGCGGCTCTAACGGAAGAAAAATCACTGCGAGTGTCTAAG	4824
Db	5016	AAATATACCGAAGAAAGATTACAGTAGACAGGAAACAAATATCACTGTAAGTATTTCTAA	5075
QY	4825	CCGAGCTCTTCAACCTTCGAGGGGATTTTTTGTCCAGGCTTAATTTCTGTGGCATC	4884

Db 5076 CCTGATGTATACCACTTCGATGTAATTTCTGACGTCCTTACATCTTTAGCAATA 5135
Qy 4885 TCCCTCCGCGATCCATGCTGATCCTGACACCTTCATTAATGATGATG 4939
Db 5136 TCATTCTTATTAATACATCATATTAATTAACAGTTTCATTTAAAAATGTAGG 5190

RESULT 7
E00656
LOCUS E00656 5760 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding p.195 protein of plasmodium falciparum.
ACCESSION E00656
VERSION E00656.1 GI:2168935
KEYWORDS JP 1986019490-A/L.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 5760)
AUTHORS Ansonli,A.H., Malkuru,J.R., Jiyasubida,S.S., Barentelina,R.M. and Karenu,G.O.
TITLE DNA CLONING FOR ANTIGEN OF PROTOZOA
JOURNAL Patent: JP 1986019490-A 1 28-JAN-1986;
COMMENT WELLCOME FOUND LTD:THE
OS Plasmodium falciparum
PN JP 1986019490-A/1
PR 22-FEB-1985 JP 1985034280
PR 22-FEB-1984 GB 84 8404692, 26-SEP-1984 GB 84 8424340 PI
ANSONLI AASBA HORDUPA, MAIKURU JIEMUSU ROTSUKUIYAA, PI
JIYASUBINA SHINKU SANDEYU, BARENTEINA RIBEROSU MORENO, PI KARENU
GERITSUTO ODEINKU
PC C12N15/00,A61K39/015,C07H21/04,C07K13/00,C12N1/20,C12P21/02,
PC (C12N1/20),
PC C12N1:19),(C12P21/02,C12N1:19);
CC strandness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH key Location/Qualifiers
FT CDS 216..5180
FT sig_peptide 219..272
FT mat_peptide 273..5177
FT Location/Qualifiers
1..5760
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
BASE COUNT 2565 a 630 c 725 g 1840 t
ORIGIN

Query Match 54.8%; Score 2706.2; DB 6; Length 5760;
Best Local Similarity 72.1%; Pred. No. 0;
Matches 3587; Conservative 0; Mismatches 1343; Indels 45; Gaps 3;

Qy 10 ATGAATATCATTTCTCTCTGTCATTCCTGTTTATCATCAATATACAGTCGTG 69
Db 216 ATGAAGATCAATATCTTTTATGTTTCATCTCTTTTATTAATAACCAAGTGTATA 275
Qy 70 ACCGACGAATCCTATACAGAGCTGTTAAAGAACTGAAGCTTTGGAAGATCCCTCTT 129
Db 276 ACACATGAAGATTAATCAAGAACTGTCAAAAACATAGAAAGCTTTAGAAAGATCAAGTATG 335
Qy 130 ACCGATACAGCTGTTTCAAGAGAGAGAAATGCTGTAATGAAGAGACAGATGGCAG 189
Db 336 ACAGTTATAGTTTATTTCAAAAGAAAAAATGTATTAATGAAGAACTGAAGACA 395
Qy 190 GCCGTACACAGCAGCAGCGGTTCTAAAGGCTGTGAGGCTAGCGGCTCCGCTGGG 249
Db 396 GCTGTACACTAGTACACCTGGTTCACAAAGGTTTCAGTTGCTTCAGGTGTTACAGTGGC 455

Qy 250 TCTGTGCGCTTGCGGGGTTCCGCTCCGCGCGCAGCGGATGATCAAGTGGCTAGT 309
Db 456 TCAATGCTTTCAGTGCGCTCACTTGTAGGTGGCTAGTGGCTTCAAGTGGCTAGT 515
Qy 310 GCAAGCGGCGTTCGCGGAACAGTGAAGAACCAATCCATCTGACACTAGCAATTC 369
Db 516 GCTTCAGGTGGTTACAGGTAATTCAGACGTAACCAATCCCTTCAGATTAATCAAGTATCA 575
Qy 370 GACGCAAGTCTTACGCCGACCTCAAGACCGAGAGAAATATCTCTCTATCAAG 429
Db 576 GATGCTTAATCTTACGCGATTTAAACACAGATGACGAATTAATCTGTTAACTATCAAA 635
Qy 430 GAGCTGAAGTACCAAGTGTGTGACCTCACTAATCATATCTGACACTGTGATATAC 489
Db 636 GAACCTCAATATCTTCAACTCTTTGATTTAATCATATCTTATCTTGTGTATAT 695
Qy 490 ATTATGCTTCAAAATATCTGATGACGTTACGAAGATCAATGAACCTCTGTACAG 549
Db 696 ATTATGCTTCAAAATATTTAATTTATGATGATGATGAAGAAATTAATGATTAATAA 755
Qy 550 TTGAATTTCTACTGACTTGTCTAAGGCGCAACTGAATGACGTTTGGCCCAATGAT 609
Db 756 TTTAACTTTTATTTGATTTATTAAGACCAAAATTAATGATGATGCTAATGATTAAT 815
Qy 610 TGTCAATTTCCATTCATTTTGAATGACAGACCAAGCTTGTGAGTATGAAGAATTG 669
Db 816 TGTCAATATCTTTTCAATCTTAAATTTGCTGCAAAATGATGACGTTATTAATAACTT 875
Qy 670 GTCTTCGATATCGCAAGCCTCTGACACAATCAAGCAATGTGGGAAGATGGAAGAT 729
Db 876 GTTTCGATATGAAGAAACCATATAGACAAATTAAGATTAATGTGAAGAAATGGAAGAT 935
Qy 730 TATATTAATAAGATTAAGAACCCATGAGAACATTAACGACCTATGAGAAATCCAA 789
Db 936 TACATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 995
Qy 790 AAGACCATAGACAAATAATAGAAATGCAACCAAGAGAGAAAGAAAGAGTGTACAG 849
Db 996 AAACCAATGATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1055
Qy 850 GCCAGTACGACCTGTGCTATATACAAACAGCTTGAAGAACCCATTAACCTATCAGC 909
Db 1056 GCTCAATATGATCTTTCTATTACAAATTAACATTAAGAAAGCAACATTAATTAATAAGC 1115
Qy 910 GTACGAGGAAGCGCATAGACACCTCAAGAAATGAATTAATTAAGAAATGCTGAC 969
Db 1116 GTTTTGAAGAAACGTAATGACCTTTAAAGAAATGAAGAAATTAAGAAATTAATGAT 1175
Qy 970 AAGATTAATGAATTAAGAAATCTCCGACCACTGTGGAAACACCCCTAACAGCTG 1029
Db 1176 AAGATTAATGAATTAATAATCCGCCACCGGCAATTTCTGGAATATACCAAAATCTCTC 1235
Qy 1030 CTGACAGAACACAGAGATAGAGAGACGAGAAAGATCAAGAGATGCCAAATCC 1089
Db 1236 CTTGATTAAGAACCAAAATAATGAGAGACACGAAAGAAATTAAGAAATTAATGAT 1295
Qy 1090 AATTAAGTCAATATGATTTCTCTTACTGATCCCTGAGCTGAGTACTACTGGA 1149
Db 1296 AATTAATTAATTAATTAATTAATTTATTTTACTGATCCATTTGAATTAATTAATTAATA 1355
Qy 1150 GAGAAGATTAAGATTAATAGATCTCCGCCAAAGTCGAGACAAAGATTAACCGAATCT 1209
Db 1356 GAAAAAATAAATAAATTAATGATTAATGTCGAAAGATTAAGAAAGATTAACGTAACCC 1415
Qy 1210 AATGAATATCCCAATGTTGTGACGATCCCTCTGTCTTAATGAATATCAACAGCTCTC 1269
Db 1416 AATGAATATCCCAATGTTGTGACGATCCCTCTGTCTTAATGAATATCAACAGCTCTTA 1475
Qy 1270 AAGCGCTCAATAGCTGTGAGTCTGATTAACCCCTGATTAAGCAAGAAACCCCT 1329
Db 1476 AATGAATTAATTTCTTTGATGATTAATTAATTCATTTGATTAATTAATTAATTAATTA 1535
Qy 1330 AAGATATCTACACAGACAAATGAGAAAGATTAATCAACGAATCAAGAGATTC 1389

D	b	1536	AAAAATCTATTACGATTAATAAGAAAATTCATTAATGAATTTAAGAAAAAAATTT	1595
O	y	1390	AAAAATGGAGAAGAAAAATTGAGGTGCACAAGAAAAAGTTACGAAGCCGACGAAAAGT	1449
D	b	1596	AAAAATGAAAAAAAATTAATGATCTGATTAATAAAATCTTGACGAAGACGATCTAAGTCT	1655
O	y	1450	CTAAACGATATCATTAAAGAGTATGAAAAAGCTGCTGACAGACATCTATGATTCCAAAATTC	1509
D	b	1656	TTTTAATGATTAACAAAAGAAATATGAAAAATTACTTAAGAAATTTATATGCAAAATTC	1715
O	y	1510	AACCATTAATCATGACCTGACCAACTTCGAGAAAAATGATGGAAAAAGCGTACTTACANA	1569
D	b	1716	AATTAATTAATATGATTTTAATCTAATTTCCGAAAAATGATGGTAATAAGATATTCATATAA	1775
O	y	1570	GTTGGAATCTGACACACCATAATATCCTTTCGATCCCTATGAGAAATTTGAAGCATATCTT	1629
D	b	1776	GTTTGAATTAATACACACCCCTAATCTTTTTCATCTTATGAAATTTCTPAACATATCTTT	1835
O	y	1630	GAGAGCTCCCAAAGCTCTTAAGTATATGAGAGACATCTCTCGCGAACATTTGTGTG	1689
D	b	1836	GAATGTTTAACAAAAGCTCTTAAATATATGAAAGATTTATCTTTAAAGAAATTTAGTAGTT	1895
O	y	1690	GAGAAAGAACTAAATATATTCAGAATCTCATATAGTAAGATGAAACAGACATCGAGAG	1749
D	b	1896	GAATTAAGAAATTTAAATATTTATTAATAATTTAATAGCAAAATGAAAAATAGATTGAACA	1955
O	y	1750	CTTGTGGAACATTTAAGAGGATGAAGAACAGTTGTTTGAAGAGAGATTCACAAAGAC	1809
D	b	1956	TTAGTGTGAAATTAATTAATAAATAGTANAGACAGCTTTTGAATTAATTTACTPAAGAC	2015
O	y	1810	GAAATTAAMCCGATGAGAGAGATCCTCGAGCTCTCCGATATTTGTTAAAGTCCAAGTCGAG	1869
D	b	2016	GAAATTAAMCCGATGAGAGAGATTTTGAAGATCTGACATTTGTAAATACAGTCTCAA	2075
O	y	1870	AAGGTGCTCCCTATGACAAAGATGTATGAACTCAAGAAAGCTCACTCTCTGGAAGAC	1929
D	b	2076	AAAGTTTATTTAATGACAAAATTTGACGATTTAAAAAAGACTCAATTTGTTTAAAAAAT	2135
O	y	1930	TGCGAGTTAAAACATATATATACATGTGCCGATAGTTATTAAGCAGAGATTAAGCAGNA	1989
D	b	2136	GTAGATTAATAACATTAATATATACATGTGCCAATCTTACAAAAGAAAAATTAAGCAGNA	2195
O	y	1990	CCATCTACTCTATCTGCTACTCAAGAAAGATAGACAAACTAAAGTGTCTATGCCCAAA	2049
D	b	2196	CCTTTATTTATTTATGTGTGTAAGAAAAATTTGATTAATTTAAAGTGTCTATGCTTAAG	2255
O	y	2050	GTCGAGAGCTGATCACGAAAGAAAGAACATTTAAACTGAAGAGACAGTCAAGTATAC	2109
D	b	2256	GTAGATCTATTATTAATTAAGAAAAAAAACATTAATAACACAAAGGTCAATCGGATTAAT	2315
O	y	2110	TCCGAGCTTCCACAGAAAGAGAGATTAACCGGACAGGCTTACCACCAAGCCCGGACAAAG	2169
D	b	2316	TGCGAAGCATCAACGGAAGAGAAATTAACAGSACAAGCAACTACAAAAACCTTGACACAGA	2375
O	y	2170	GCCGCTTCACTCTCGAAGGCGATAGCGTGAAGTCAAGCAACAAGAGCAGAGAAGCA	2229
D	b	2376	GCAGATCTGCTTTGTAAGAGAGATTCAGTACAAAGCAACAAGAACAAAAACAAGA	2435
O	y	2230	CAGCCTCCAGTCCAGTGGCCCTTCCAGAGGCTAAAGCTCAAGTGCCTACACCCAGCT	2289
D	b	2436	CAACCAACAGTACAGCTATACAGTACCAAGCAAAAGCAACAGTCCCAACACACACAGA	2495
O	y	2290	CCTGTGATTAACAAACGAGAAATGTACAGCAACTGAGTACTTCTGAGAGAGCTATGAG	2349
D	b	2496	CCAGTAATTAATTAATACTGAAAAATGTTTCCAAATTAAGATTATCTTGAAAAAATTAATCA	2555
O	y	2350	TTTCTGAATATACCTCTACATCTGCACAATAATATATCTCTCTCAGCAGCTATGAAC	2409
D	b	2556	TTTTTTAATATCTTCAATATATATGTCACAATAATATTTTGGTTTCAACACTCACTATGAAC	2615
O	y	2410	GAGAGATTTCTTAACAGTACAGATTAACCAAGAAAGAGAGAGATTAAGTCTCTTGT	2469
D	b	2616	GAATGATTAATTAACCAATTAATAATTTCAAAAGGAGAAAGCAAAATTAAGTCAAGT	2675
O	y	2470	GATCCACTGAGACTGCTGTTCAATATCCAGAAACAACATTTCCGTTATGATTTCTATGTTCC	2529
D	b	2676	GATCCATTAGACTTATTTGTTTAATTAATTAACAAAATACATCGTATATCTATCTATGTTT	2735
O	y	2530	GATAGCTCAACA-----ATTCCT	2547
D	b	2736	GATAGCTTAACATTAAGTTTATCAACAATTTTATGGGTTTATGAAAAAGAAATGEGTTAT	2795
O	y	2548	CTCTCTCAACTGTTTCAGATTAATATAGAAAGAGATGTTCTGCACACTGTATTAACCTC	2607
D	b	2796	GTATTTAATATCTTATGTAATTTATGAAAAAGAAATGGTTTGAATTTATATTAACCTT	2855
O	y	2608	AAAGCAACGACAGATTTAGAACCTTCTGAGAGAGCTAAGAGGCTCCACCTCTGTT	2667
D	b	2856	AAAGTAATTAATGCAATTTAAAAATTAATTAAGAGAGGAAAAAAGTATCCACATCTTA	2915
O	y	2668	AAACTCTCTCTCCAGCTCCATGCAACACAGTCTGCACACCTCAAGACAGGCGGA	2727
D	b	2916	AAACTCTTCAAGTTCAATCAATGCAACCATTTATATTAACCTCCAGATTAACCCGAA	2975
O	y	2728	GTAGCGCTTAACGACGACACCTCTCACTGACCAACTTAATMACTCACTGAATCTGTT	2787
D	b	2976	GTAACTGCAATAGATGATATACATCACTCTCAATTAATTTGAATTAATGTTTAAATTAATTT	3035
O	y	2788	GAGAACATCTGCTCTCTGCGAAGATTAAGAACATCTCAACAACTTAATTTGACAGAAA	2847
D	b	3036	GAATTAATTAATGAGCTCTGGAATTAATAATAATAATTAATTAATTAATTAATTAATTAAT	3095
O	y	2848	TGCTCCGAACTCTTACGAGAGATTAACGAAAGACGACGACACTTCTATPACGAGAGC	2907
D	b	3096	AGTACTGAATCTTTTAATGAAGATTAATAAGATGATGATCTTTTATTAATGAATCT	3155
O	y	2908	TTCACTTAATCTGTAATTTAAACCCGATATCACTCTCTTAACGATTAATCTAAA	2967
D	b	3156	TTTACAATTTTGTAAATTAATTAACCTGATGATTAATTAATTAATTAATTAATTAATTAAT	3215
O	y	2968	CGTAAGACCTGGAAGAGACATCAATAGCTGAAGAAAGACCTGCAACTGAGCTTGAC	3027
D	b	3216	AGGAAGAAATTAAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3275
O	y	3028	CTGTACACAGTACAACCTGAACTGAGAGACTCTCTGCAAGAAAGACAGCTGGC	3087
D	b	3276	TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3335
O	y	3088	AAATTAAGTCAAGTCAAGAAAGTGTCTGCTCAAGAGACAGCTGGAAGCAACATC	3147
D	b	3336	AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3395
O	y	3148	AACTCACTGAACCAATCGAAGACAGCTACT	

Oy	4587	CATCTCCAGCAGCCAAATGGGTGAAGAAACAGTG--CCCCGAAATAGGGCGTGTTCAGG	4644
Db	4836	CATTTACACACCCCAATGGGTATAAAAACAATGATCCACAGAGTTGTGGATGTTTTCAGA	4895
Oy	4645	CATCTGGAGCAGCGCGGAAGAGTGCATGAGTCTCCTGAACCTACAAACAAGAAGAGATTAAG	4704
Db	4896	CATTTAGATGAAGAAGGAAGAAATGTAAATGTTATTATTAATTCAAACAGAAAGGTAGTAA	4955
Oy	4705	TGCGGGAAGAACCCAAACCTTACCTCGCAATGAAGAAACATAGCGGGGTGAGCGCGATCT	4764
Db	4956	TGTGTTGAATATTCAAATCTACTTGTGTAACGAAATTAATGGTGGATGGATGACAGATCC	5015
Oy	4765	AAATCGACCGAGGAAGACAGCGGCTCTAAACGAAAGAAATAATCACATGCGAGTGTACTAAG	4824
Db	5016	AAATGTACCGAAGAGATTCAGTAGTAGCAACGAAAGAAATACATCATGTGTAATGTACTAAA	5075
Oy	4825	CCCGACCTCTATCCACTCTTGACGGGGATTTTGTGCTGCACGATCAATATTCCTGGGCGCAT	4884
Db	5076	CTTGATGTTTATCCACTTTTCATGATGTAATTTTCTGCAAGTTCCTCTAATCTTCTTAGGAATA	5135
Oy	4885	TCCTTCTGCGTGATCCTCATGCTGATCCTGTACAGCTTCATCTAATATAGATGATG	4939
Db	5136	TCATCTTAATTAATCTCATGTTATTAATTAACATCTTCATTTAAAAAATAGAG	5190

RESULT 8	
PEP190	
LOCUS	5282 bp DNA linear INV 21-FEB-1996
DEFINITION	P.falciparum gp190 (MSA1, MSP1, PMMSA) gene for merozoite surface antigens precursor.
ACCESSION	X03371
VERSION	X03371.1 GI:929797
KEYWORDS	gp190 gene; merozoite surface antigens; MSA1 gene; MSP1 gene; PMMSA gene.
SOURCE	Plasmodium falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 5282) Mackay, M., Goman, M., Bone, N., Hyde, J.E., Saito, J., Cerra, U., Stuenkelberg, H. and Bujard, H.
TITLE	Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level
JOURNAL	EMBO J. 4 (13B), 3823-3829 (1985)
MEDLINE	86136024
PUBMED	3004972
REFERENCE	2 (bases 1 to 5282) Pan, W., Tolle, R. and Bujard, H.
AUTHORS	A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1
TITLE	Mol. Biochem. Parasitol. 73 (1-2), 241-244 (1995)
JOURNAL	96123395
MEDLINE	8577332
PUBMED	3 (bases 1 to 5282) Bujard, H.
REFERENCE	Direct Submission
AUTHORS	Submitted (23-JUN-1995) H. Bujard, Zentrum fuer Molekulare Biologie der Universitaet (ZMBH), Im Neuenheimer Feld 282, Heidelberg 69120, FRG
TITLE	On Aug 5, 1995 this sequence version replaced gi:9920.
JOURNAL	Location/Qualifiers
COMMENT	1..5282
FEATURES	/organism="Plasmodium falciparum"
source	/isolate="K1"
	/db_xref="taxon:5833"
	/chromosome="9"
	/dev_stage="blood stage"
	72..4964
gene	/gene="gp190, MSA1, MSP1, PMMSA"
	72..4964
gene	/gene="gp190, MSA1, MSP1, PMMSA"
	/codon_start=1
CDS	/product="precursor for the major merozoite surface antigens"

protein_id="CAA27070.1"
/db_xref="GI:329798"
/db_xref="SWISS-PROT: P04932"
translation="MKILPFLCSLEPFLINTQCVTHESIQEQLKRLLEDAVLTGYVS
LFLHKKMLNEELTITLGAASQSGTSGTSSQPSQSTSSSKSNLTLPSSNSGAG
SPEADASQSDKSTADKLKHVRYNLITLIELKQPLDTLMTLNLCHNGFYRLID
GYEENELLYKLYNFGPLLKATLVNCADVCOIPFLNLTANEDLVYKLTCPFGYRKR
LDNTRKQVGMEDYITKERTKININIELDESKTIDKTRKTPANEKKEKTLQVAYO
SLYKQLLEAHNLTIVLEKTRIDPLKNNENITLKLKININIKRPPANSKNTPLNDR
MKILKEHKEKLELAKTRIKNIDSLFDTLPDLYEYRERKNIDTSAVEYRKESEPR
ETKQVTLPLSDKIDINAMLESLSGFLNIDPDTYKREPKNITDNEKKITNEKEK
ITIEKRTIEKSTQYEDRSWLDITREYKRLINEYDSRFPNNIDTLTEFKMGKRTK
SYEKETLHNTHTPVSYSNKHKEITKRLKATWEDSLNIVOEKELRYKYNILSKLE
NEITVELVNIKKDEQLEQSHKRIKDEKNDPEKLTLEVDLIVKVOALMLKIDELK
TOLLKANEKLNHNVSQKONKQEPYLIYKLEIDRLKVPYRYESILNEEKKRYL
ITIEQSDQSNSESPGEGITQCATQTPQOAGSLBEDSQAOAOCQAOOPVAPVAP
EAKQVPPAPVAPKNTKENTIDYETLKYPLMSTYSICHTYILVSHSTNMTILKQY
KITREESKLESCQDPLENNIONTIYKMLSPMSYINSLTQIMPEYTERKMYCNTRYK
LKNDKRLKMLEEKKKSTVSTKLSSSMQPSLTLPDQKREVSANDJSHSTLNANSI
KLFEVTSLEGNNKIYELLGQKSNSEYFKLTKSDSYVNSFPLNYSFVSKRADTINSI
NDESKRRLLEEDINKRLTILQSFDTLKLLEFLPEKRTVAKYKMQKILKTLTKL
BOLESKRLSLNPNHVSQNSVFNFKKKEALIEENLENTLENTLIRKLYKGLIVLYN
ESSPLKTLSEESIETDENVASILENFKYLSLEKLEKLDNLEKTLKATYSLGSLYK
ELKEVYKRNKRTQSPENNTDNNALDEYKRLPGCTGVAYVSESSQDTLEOSPQD
LKPSHVGAESENITTSQSNVDVDAEYKRLTPEGESEYDVLGGVYGAATVYSVID
NLSITNEEYVYKILKALVAYSYSKLOLNNMNPVAVKIDNLSRENKREKNTVYK
ESDLTPKQDLSSNYVYKDKYKRLNEKREKRLSTSYIKSDITPDIINADULCYIKY
ILSEYTSQSDLSIKRTINDQGEKREKLEPPLNITELYLVTVNDKIDLYETLKEAVL
LYTESKNEVYIKELANTLQKILSOQKADPKRNNNVEIADLSDVYNNHLITKPLSTG
VYENIAKATVSLNLDGKLOGLNILOHQCYKQPDQSGCFPHLDEDEKCLNLYKQ
GQKQVKNPENTCENNNGQDAQKCTEEDSGSNGKRTICECTKPDSPPLDGDGIFCS
NFIPLISFLITMLILNLSPI"

sig_peptide	72	.128	/gene="gp190, MSA1, MSP1, PMMSA"
			/note="putative"
BASE COUNT	2363	a	608 c 676 g 1635 t
ORIGIN			

Query Match	54.4%	Score 2686.8	DB 3	Length 5282
Best Local Similarity	71.8%	Pred. No. 0		
Matches 3541: Conservative	0	Mismatches 1362	Indels 27	Gaps 1

Oy	10	ATGAATATCATTTCTCCCTGTCATCTTCTGTTTTATCATCAATACAGTGGCG	69
Db	72	ATGAAGATCATATTCCTTTTATGTCATCTCTTTTTTATATATAATACAAATGTGTA	131
Oy	70	ACCCACGAATCCTATCAGAGAGCTGTTAAGAACTGGAAAGCTTTGGAGATCCGCTCT	129
Db	132	ACACATGAAAGTTATCAAGAACTTCTCAAAAAACTAGAAGCTTTGAAATCAGATATG	191
Oy	130	ACCGATACAGCTCTTCGACGAAGAGAAAGTGGCTGAATGAAGAGGACGAGTGGCAG	189
Db	132	ACAGGTTATAGTTTATTTTCATATAGGAAAAAATGATCTTAAATGAAAGAAATTT	245
Oy	130	GCCGTTACACACGACACCCCGGTTCTTAAGGCTGTGTCGTAAGCGGTGCTCCGCTGG	249
Db	246	-----ACTACAAAGGTGCAAGTGTCAAAAGTGTCAAGTGT	284
Oy	250	TCTGTGGCCTCGGGGGTCCCTGCGCTCCGGCGAGGAGTGACATCAGGTGAGTCAGTG	309
Db	265	ACAAGTGTGACAAAGTGTGCCAAGTGTGCCAAGTGTGACAAAGTGCATCTGCTTCAAC	344
Oy	310	GCAAGCGGCGGTTCCGGGGAACAGTGGAAAGAACCAATCCATCTGACAACTTAGCGATTCC	369
Db	345	ACTTTACCTCGCTTCAAAATCTTCATCTGCTGCAAGCCCTCCAGCTGATCAAGACGATTCA	404
Oy	370	GACGGCAGTCTTACGGCGAGCTCAAGACCGAGAGAGAACTATCTCTCCTACTATCAAG	429
Db	405	GATGCTTAATCTTCAGCTGATTTAAACACAGAGTGCAGAAATTTACTTGTAACTATCAAA	464
Oy	430	GAGCTGAAGTACCCACAGTGTGTGCACTCACTAATCATATCTGTACACTGTGTGATAAC	489
Db	465	GAATCCAAATATCTTCAACTCTTGTATTTAACTATCATATCTTAACTTGTGTGTATAT	524

QY	430	ATATGCGCTCAAAATATCGATTGACGGTTACGAGAGCATGAACTCCTGTACAG	549
Db	525	ATTCATGGTTCCAAATATTTAAATGATGATATGAGAAATTAATCAATATATATATATAA	584
QY	550	TTGAATTTCTACTCGACTGCTAAGGGCCAAACGAAATGACGTTGGCCAAATGATAT	609
Db	585	TTAAACCTTTATTTTGTATTTATTAAGAGCAAAATTTAAATGATGATGTGCTATATATAT	644
QY	610	TGTCAAATTCATTCATTTTGAAAGATCGAGCCACAGAGTGGAGCTATTTGAAGAAGTTG	669
Db	645	TGTCAAATACCTTTTCATCTTAAATTCGTCGCAAAATGAAATGAGCTACTTAAAAACGT	704
QY	670	GCTTCGGATATCGCAAGCCTCTCGACACATCAAGACATGTGGGAAAGATGACAGAT	729
Db	705	GCTTGCGATATGAGAAACCATTAAGACAAATTTAAAGATTAATGTGAGAAATATGAGAAAT	764
QY	730	TATATTTAAAAAGATTAAGAGACCATTCGAGACATTTAACGAGCTGATCGCAAAATCCAA	789
Db	765	TACATTTAAAAAAATTTAAAAAACCATTAAGAAATTTAAATGAATTAATTTAAACAAAGTAA	824
QY	790	AAGACCATAGCAAAATTAAGAAATGCAACACAGAGAGAGAAAGAAAGAAAGTTGTACAG	849
Db	825	AAACAAATTTGATTAATAATTAAGATGCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	884
QY	850	GCCCAAGTACGACCTGTCCATCTATTAACAAACAGCTTGAAGAGAGCCCATACCTCATACG	909
Db	885	GTCATATATGATCTTTCTATTTACAAATTAACATTTTGAAGAGACATATTTATTAATAGC	944
QY	910	GTACTGGAGAGGCGATAGACCCCTCAGAGAAATGAAATTTCAAGAACTGCTGCAC	969
Db	945	GTTTATGAGAAACGATTTGACACTTTAAAAAAATGAAACCTTTAAGGAATTTACTTGAT	1004
QY	970	AAGATTAATGAATTAAGAAATCCGCGCCGACCCACACTGTGGGAAACCCCTTAACGCGTG	1029
Db	1005	AAGATTAATGAATTTAAAAATCCGCCACGCGCCAAATCTTGGAATATCCACAAATACTCTC	1066
QY	1030	CTGCACAGAAACAAGATATAGAGGACACAGAGAAAGAGATCAAGAGATCGCCAAACCC	1085
Db	1065	CTTGATTAAGACAAAAAAATCGAGAGACAGGAAAGAAATTAAGAAATTTGCCAAAACCT	1122
QY	1090	ATTAAGTTCAACATAGATTTCTCTCTTACTGATCCCTTGAGCTGGAATGACTTGTGGA	1144
Db	1125	ATTTAATTTAATTTGATGATTTATTTAGCTCACTGTAATTAAGAAATTAATTTAAGA	1188
QY	1150	GAGAGATTAAGATATAGACATCTCCGCAAGCTGAGACAAAGAGATCAACCCGACCT	1209
Db	1165	CAAAAAATTAATAATTTATGATATTAAGTCGCAAGAGTTGAAACAAAGAAATCACTGACCC	1244
QY	1210	AATGATATCCCAATGCTGTGACGTACCCCTGTCTTATTAAGATATCAACAACGCTCTC	1265
Db	1245	AATGATATTCCAATGAGGATTAATTCCTTTGTCATATTAACGATATTTAAACATGCTTTA	1304
QY	1270	AACGAGCTCACTACCTCGGTGACTTATTTAACCCCTTGATATATCGAAAGAACCCCTCT	1322
Db	1305	AATGAACTTAATCTTTTGGATTTAATTAATTCATTTGATTTATVCAAAAGAACCAAGT	1366
QY	1330	AAGATATCTACAGACATAGAGAGAAAGAGTTATCAACAATCAAGAGAGAGATC	1388
Db	1365	AAAAACATATTTACTGATTAATGAGAGAAAAAATTCATTAATTAATAATTTAAGGAAAAAAT	1422
QY	1390	AAATTTGAAGAGAAATTTGAGAGTGCACAGAGAAAGTTTCGAGAGAGCGCAGCAAAAGT	1445
Db	1425	AAAAATGAGAAAAAATTTGATCTGATTAATAAAAAATCTTACAGACAGATCTAAGCT	1488
QY	1450	CTAAACGATATCTAAGAGATAGAAAGCTGCTGAGACAGATCTATGATTTCCAAATTC	1505
Db	1485	TTAATGATATTAACAAAAAGATATGAAAAATTTACTTAATGAATTTATGATGCAAAATTC	1544
QY	1510	AAGATATACATGACCTGACCAACTTCGAGAAATATGATGGGAAAGAGATCTTACAAA	1566
Db	1545	AATATATATATATGATTAACTAATTTTGAAAAATGATGGTAAAGATATTCATATTAATA	1604

OY	1570	GTGAGAACTGACACACCCATTAATACCTTTGCATCTCATGACATTTCTAAGCATATATCTT	1629
Db	1605	GTGAGAACTTACACACCCATTAATACCTTTTGGATCTCTAAGAAAATTTCTTAACCATATCTT	1664
OY	1630	GAGAGCTCACCAAGCTCTTAAGTATATGAGAGGACATATCTCTCCGGACACATTTGGTGTG	1689
Db	1665	GAAAAGTTAAACAAAGCTCTTAAATATATGGAAGATATATCTTTAAGGAATATATGATGTT	1724
OY	1690	GAGAAAGACTTAAAGTATTTCAAGAATCTCATTAAGATGCAAAAGCAAGATGAGACG	1749
Db	1725	GAAAAGATTTAAATATATTTATAAAATTTAATAGCAAAATAGAAAATAGATTTGGAACA	1784
OY	1750	CTTGTTGAGAACATTTAAGAAGGTGAAGAACAGCTGGTTTGAGAGACAGATTACAAAAGC	1809
Db	1785	TTAAGTAAATATTTAAAAAGATGAAGAACAGCTTTTTGAAAAAAAATTTACTAAAGAC	1844
OY	1810	GAAATTAACCAAGTGAAGATCTCGAGAGCTCCGATATTTGTAAAGTCCCAAGTCAG	1869
Db	1845	GAAATTAACCAAGTGAAGATTTTGAAGTATCTGCAATTTGTAAAGTACAAAGTCTAA	1904
OY	1870	AAGGTGCTCATGAACAGATTTGATGAAGTCAAGAAAGCTCAACATCTATCTGAGAAC	1929
Db	1905	AAAGTTTATTAATGAACAAATTTGACGAATTTAAAAAGCTCAATGATTTTAAAAAT	1964
OY	1930	GTGAGTTAAAACTAATATATCACTGTGCCAATAGTTATAGCAGACAGATTAAGCAGAA	1989
Db	1965	GTAAATTTAAACATTAATATATCACTGTGCCAATTTCTTACAAAAGAAAATTAAGCAAGAA	2024
OY	1990	CCATACACCTCATGCACTGACAGAAAGATTAAGCAAACTGAAGTCTCATGCCCAA	2049
Db	2025	CTTATTTATTTAATGTGTGTTGAAAAAGAAATTTGAATTTAAAGTCTTATGCTTAAG	2084
OY	2050	GTCGAGAGCCTGATCAACGAAGAGAAGAAACATTTAAACTGAAGACAGCTCAGATTAAC	2109
Db	2085	GTAAATATCATTTAATTAATGAAGAAAAAAAACATTAACAGAAAGTCATTCGGATATAT	2144
OY	2110	TTCCAGCCTTCCACAGAGAAGATTAACCGGACAGGCTACCAACAGCCCGCAACACAG	2169
Db	2145	TCGGAACCATATCAACCGAAGGAGAAATTAACAGGACAAAGCAAGCACTTCCGACAAACAA	2204
OY	2170	GCCGGTTCAGCTCCGAAGGCGATACGTCGAAGCTCAAGCAAGCAAGACAGACAGACAGCA	2229
Db	2205	GCAGGATCTGCTTTAGAGGAGATTTAAGTACAGCAAGCAAGCAAGCAAGCAAGCAAGCA	2264
OY	2230	CAGCCTCAGTGCAGAGTGCCTGTTCCAGAGGCTAAAGCTCAAGTGCCTCAACACACAGCT	2289
Db	2265	CAACACACAGACAGTACAGTACCAAGAGAAGCAAGCAAGCAAGTCCCAACACACACAGCA	2324
OY	2290	CCTGTGAATTAACAGACCGAGAAATGTCCAGCAAACTGACATACCTTGAGACGCTCATGAG	2349
Db	2325	CCATTAATATATAACTGAAAATGTGTTCCAAATTAAGATATCTTGAAAAATTTATATGAAA	2384
OY	2350	TTTCTGAATACATCTTACATCTCCGCAAAATTAATCTGTCTGTCCAGACAGATATGAC	2409
Db	2385	TTTTTAATATCTTCATATATATATCTCAAAATATATTTGGTTTCCACTCAACATATGAC	2444
OY	2410	GAGAGATTTCTTAAACAGTACAGATTAACAGAGAAGAGAGATTAACCTGCTCTTGT	2469
Db	2445	GAAAAGATTTAAACATATTAATTAATTAACAAAGAGAGAAAGCAAAATTAAGTTCATGT	2504
OY	2470	GATCACTGAGCCTGCTGTTCAATTAATCAGAAACAACTTCCGTTATGTATTTATGTTTC	2529
Db	2505	GATCACTTAAGACCTTAATGTTTAAATATACAAAATTAATACATACGTAAATGATTTATGTTT	2564
OY	2530	GATGACCTCAACAACTTCTCTCTCAACTGTTCATGAGATATATGACAGAGAGATGTC	2589
Db	2565	GATGACCTTAACAACTAATTTATACAACTATTTATGGAATTTATGAAGAAAAGAAATGTT	2624
OY	2590	TGCACACGTATTAACCAAGAGCAAGCAAGCAAGATTAAGAACCTCTGAGAGAGATGAG	2649
Db	2625	TGTATATTTATATTAACCTTAAGGATATGCAAAAATTTAAAAAATTTATTAAGAGAAAGCGAAA	2684
OY	2650	AAGGTCTCACACTCTGTGTTAAACATCTCTCTCTCAGCTCATCAACCACTGCTCTCA	2709

Db	2685	AAAGATATCACATCTGTGTAAGAAACCTCTTCAAGTTTCATCATATCCAAACCATTAATCAATTAACA	2744
Qy	2710	CCTCAAGACCAAGCCCGAAGTAGAGGGCTAACGACGACACCTCTCAGTCGACCAACTTAT	2765
Db	2745	CCTCAGGATTAACCCCGAAGTAGAGGCAATAGATGATACATCTTCACAAATTTTGAT	2804
Qy	2770	AACTCACTGAACACTGTTTGGAAACATCTGCTCTCGCGAAGAAATAGAAACATCTACCA	2829
Db	2805	AATAGTTTAAATATTATTGAAAACATATTGAGCTTGGAAAAACAAAATATTATACCA	2864
Qy	2830	GAACTTATTGGACGAATATGCTCCGAGACTTACGAGAAATACTGAAAGACGCGAC	2889
Db	2865	GAAATTAATAGGTCAAAAAAGTAGTGAACACTTTTATGAAAAATATTAAATATGTAT	2924
Qy	2890	ACATCTTAACGAGAGCTTCAGTACACTTCTGTAATCTAAAGCCGATGATTCACCT	2949
Db	2925	ACATTTTATTAATGATCTTTTACAAATTTTGTAAAACTTAAGCTGATGATATTATTTCA	2964
Qy	2950	CTTAACGATGATCTTAACGATAGAAAGCTGGAAGAGGACATCAATTAAGCTGAAGACA	3009
Db	2985	TTGATATGATGATCAAAAAAGAAATTTAGAAAGAGATATTATTAATTAATAAAAACT	3044
Qy	3010	CTGCACTGAGCTTCGACCTGTACACACAGTACAAACTGAACTGGAGAGACTCTTGAC	3069
Db	3045	TTACGTTATCATTTTGGATTTTATATTAATTAATTAATTAATTAATTAATTAATTAAT	3104
Qy	3070	AAGAAGACGAGTCGCGACGATTAAGATACATCAACAGCTGACCTCTGCTCAAGAG	3129
Db	3105	AAAAAGAAAAAGTTGGTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3164
Qy	3130	CAGCTTGAAGAACATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA	3189
Db	3165	CAATTAGAAATCAAAATTTGAATTCCTTAATTAACCCAAAGCATGTATTACAAAATTTCT	3224
Qy	3190	GTGTTCTTCAACAAAGAAAGGAGGAGATGCGCGACAGAGAACACTTGGAGAC	3249
Db	3225	GTTTTCTTTAACAAAAAAGAAAGCTGAATAGCAGAAACGTAACCATTAACAAAC	3284
Qy	3250	ACCAAGATTCTTCTCAACCACTCAACAAAGGCTCTGCTACATTAATTAATGAGCAGCTCT	3309
Db	3285	ACAAAATATTATTGAAACATTTAAAGCACTTGTAAATATTATTAATGATGATCATCT	3344
Qy	3310	CCTCTGAAGACTCTCTCCGAGAGAGCATCCAGCCGAGATTAAGTACGCCAGCTCGAG	3369
Db	3345	CCATTAAAACTTTAAGTGAAGATCAATTTCAACAGAAAGATTAATTAATGACGTTAGAA	3404
Qy	3370	AACTTCAAGTCTCTGTAAAGCTGGAAGGCAAGCTAAGACAAACCTGAACTGGAGAG	3429
Db	3405	AACTTTAAAGTATTAGTAAATTAAGAAAGAAATTAAGAGATTAATTAATTAAGAAAG	3464
Qy	3430	AAGAAGCTCAGCTCTCTCTAGCGGAGCTGATCACTATGCGCGAGCTCAAGAGACTC	3489
Db	3465	AAAAATATTATCATCTTATATCAAGTGGATTACATCAATTTAATGCTGAATTAAGACAGTA	3524
Qy	3490	ATTAGAACAAGAACTACACCGGCAATAGCCCAAGCGAAGATTAATACAGACTGATTAAC	3549
Db	3525	ATAAAAAATAAATTTATACAGGTAATTTCCAAAGTGAAGAAATTAATACGAGATTTAACAT	3584
Qy	3550	GCACTGAATCTTACAGAAGTTCTGCTGGAAGAACAGATGTGCCACTGTGTGTCT	3609
Db	3585	GCATTAGATCTTTACAAAAATTTCTCCGAGAGAACAGATGTTGCAACAGTTGTAGT	3644
Qy	3610	GAACTGGCTCCGACACACGAGGAGCACTGCACACTTAAGAACCTGATCACTACATCTC	3669
Db	3645	GAAAGTGAATCCGACCATTAAGAACAAAGTCAACCAAAAGAACAGACTCAACTCATGTA	3704
Qy	3670	GGAGCCGAGTCAATTAATTAACATATCTCAAGAGTGGAGAGTGAAGTGCATGATGCTC	3729
Db	3705	GGAGCAGAGTCTTAACACATTAACACACATCAACAAATGTGATGATGAAGTATGACGTA	3764
Qy	3730	ATCATTTGGCTATCTTGGCGAGAGCGAGAGGAGACTACGATGACCTGGCCAGGTGCTC	3789

OY	1972	CAGGAAATTAAGCAGGAACCATCTACTACCTCATCGTGTACTACAAGAAGATAGCAAACTG	2031
Db	2043	CCACAAATTAACCCAGAACCATATTATTTTAATTGTAATTAATAAAAAAGAGTGATTAATTA	2102
OY	2032	AAAGTTTCATGCCCAAAGTCGAGACCCTGCATCAACGAAGAACAAACATTAATAACT	2091
Db	2103	AAAGATTTTATACCAAAAAGTAAGAAAGCATGTGTTAAAGAAAGAACAMGGCTGTTATCAAGT	2162
OY	2092	GAAGGACAGTC-----AGATACTCCGAGCCTTCACAGAAAGGAGAGATTAACGGGACG	2145
Db	2163	ATTACACAACCTTTTAGTTGGAGCAACCGAAACCACTGGAAGATGGGGGTCACTCCACACAC	2222
OY	2146	GCTACACCAAGCCCCGAGCAACAGGCGGTGCTGCTCGAAGGGATAGCGTGCAGCT	2205
Db	2223	ACATTATTCCAATCAGSAGAAACAGAAAGTAACGAGAAACGAAACGAAACGAAACGAAACMA	2282
OY	2206	CAAGCACAGAGCAGANGCAGG-----ACAGCCTCGAGTCCAGTCCGCGTTCCA	2256
Db	2283	GTAGGACACMACMACMGTTAACATPACTATTCACCCACCAACACATCAACACACAAA	2342
OY	2287	GAGGCTAAAGCTCAAGTGCCTACACCACCGCTCGTGAATPAACAGCCGAGATGC	2316
Db	2343	GAGTAAAAAGTTGTTGAAATTTCAATATAGAACATAGAGATATGACATATTCACAAAGCTTG	2402
OY	2317	AGCAAACTGCATCACTTGAGAAAGCTGTATAGTCCGTGAATPACATCCTPACATCTGCCAC	2376
Db	2403	ACAAAACAGTTTATTCATAAGAAATATAGATGAATTTTTTAACTAAATCAATATATGTCAT	2462
OY	2377	AAATATATCCGCTCTGCACACACATAGAACAGAAATCTTAAACAGACAGATTA	2436
Db	2463	AAATATATTTTAGTATCAAACTAGATATGACAAATAATTTTAGAGATATATATCTT	2522
OY	2437	ACCAAGAGAGAGAGACTAAAGTCCTCTGTGATCACTGGACCTCGTTCATATTC	2496
Db	2523	ACTCCAGAGAAAAAATGATTAATAATCATGTGATCCATTTGATTTTATTTTAAATAT	2582
OY	2497	CAGAAACAATCCCCTATATGATTCATAGTTGATAGAGCTCAACAAATTCCTCTGCTCA	2556
Db	2583	CAAAATATCACTCTGCTATATGTTATTCATATATAGATATGATTAACAAATGATTTTACACAT	2642
OY	2557	CTGTTCAATGAGATATATGAGAGAGATGGTGTGCACACTGTATAACTCAAGACAC	2616
Db	2643	CTCTTTTGTGAATTAATCAAAAGSAAAGATTTATTTATTCATTAACCTAAAGAGGAA	2702
OY	2617	GACAAGATTAAAGACCTTCGGAGGAAGCTAAGAGCTC-----	2657
Db	2703	AATCACTCAAAAAATTAATTAGAGGACAAAAACAATATACGTGAACATCATTAACCTC	2762
OY	2658	-----CACCTCTGTAAACCTCTCTCCAGCTC	2687
Db	2763	AGTCCTGGAATPACAAACCGTAATTAATCTGGCTCAATCCGCAATCAAGTAATTTCCCAAAC	2822
OY	2688	CATGCACACACTGTCTCTACACCTCAAGCAA-----G	2721
Db	2823	CAACAAATCAATGATCTCTTACCATTACCCAAATAGGTGATGATATCTGATCT	2882
OY	2722	CCCGAATGAGCGGTATACGACGACACCTCCACTGCACCAACTTAATATCCAGTAA	2781
Db	2883	GCTGTAGTTGAAGAAATCTATGATCCCTTACACAGTATTTGCTATTTAGTATGACGATTTGAA	2942
OY	2782	CTGTTTNGAACAATCCCTGTCTCTCGGCAAGAAATATAGAACAT---CTACCAAGAACTTAT	2838
Db	2943	GGTATTTGTTAGTCTCTTAATTCCTTGAATAATAAACCTAAAGTACCTATTCATTAACCAT	3002
OY	2839	GGACAGAAATCTCCGAGAACTTCTACAGAGAGATATCTGAAGACAGCGACACATTCAT	2898
Db	3003	TCTACACAGAGATGAGAAAAATTTTATGAGAAATATTTTAAAAAATAAGATACCTATTTT	3062
OY	2899	AACGAGGCTTCACTACTCTCGGAAGAAATCAAAAGCCGATGATATCACTCTTAACGAT	2958
Db	3063	AATGATGATATCAACCAATTCGTAATATCAATATCAAAAGTAATTAACAGGTTTGAC---	3119

[illegible]

Db 4197 TCTGATTTAATGCAATTTAAACATATATCCCAATGAAATGATATATGGAATTCATTT 4256
Oy 4060 AAGTTCCTCAATTAAGACAGAGGATTAATTTCTGCTAGTACACTATATCAAGAC 4119
Db 4257 AAAATATATGATTTAGAAACAAAACAAACACTTTTAAAAATGTTAAATATAAAGAA 4316
Oy 4120 TCCATGCGACCGGATATCAATTTGCTTAATGATGCTGGGTATTTACAGTCTGAGC 4179
Db 4317 TCAGTAGAAATGATATTAATTTGACAGAGGATATAGTATTAAGAAAAGGTTTAA 4376
Oy 4180 GAAAAATCAAGTCTGACCTGACTCTATTAATAAAGTATTAACAGATAG----- 4230
Db 4377 GCGAAATTAAGATGATTTAGATTTCAATTAATAAAGTATTAAGAAAAGAGGTTTC 4436
Oy 4231 -----CAAGCGAGAT 4242
Db 4437 CCATCATCACCACCAACACCTCCGTCACGCAAAAAGACGAAACAAAGAGAA 4496
Oy 4243 GAAAAATATCGCCCTTCCGTAATACATGCAAAACCCGTCAAGACAGTCAACGACAA 4302
Db 4497 AGTAAGTTCCTTCAATTTTAAACAAACATGAGACCTTATCAATTAATGTTAATAA 4556
Oy 4303 ATGACCTCTCTGTAATTCACCTGAGGCAAGGTCTCACTACTACTTACGAGAGAGC 4362
Db 4557 ATGACGATTAATTAATTAATTAAGGCAAAAGATTAACGATTTGATGTTGAAAAGAT 4616
Oy 4363 AATGGAAGTAAATCAAGAGCTGAACTCACTCAAAACATCAAGACAGAGCTGCA 4422
Db 4617 GAAGCAGATGTTAAATTAAGTAACTAGTATTAAGCAATGATGACAAATAGAT 4676
Oy 4423 GATTTCAGAAAATTAACAATTTGCTGCGAATGAGACCTGCTACCGATTTATACAC 4482
Db 4677 CTTTTTAAACACCTGACCTGCAACCAATTAATAATGATATGATGATGAGAA 4736
Oy 4483 AACATCTCTGACCAAGTTTCTGTCACCTGAGCTGTTGCGAAGACCTGCCAAAACA 4542
Db 4737 AAAGATATGCTGGCAATTTACTTAGTACAGAGATTAG---TTCAAAATTTTCTTAACA 4793
Oy 4543 GTGCTGAGCAATTCGCTGAGGCAAGCTGCAAGGCTGCTGAACATCTCCAGCACCA 4602
Db 4794 ATAAATCAAAATTAATTAAGGAAATTCCAAGATGTTAAACATTTTCACACACCA 4853
Oy 4603 TGGCGAAGAAACAGTGGCCGAGATAGCGCTGTTTCAGGATCTGAGACGAGCGCA 4662
Db 4854 TGGCAAAAACCAATGTCGCAAAATTTCTGATTTTCAGACATTTTATGATGAAGAGA 4913
Oy 4663 GAGTGCAGTGTCTCTGACCTAACAACAGAGAGATTAAGTGGTGGAGAACCAAC 4722
Db 4914 GAATGTAATGTTTATTAATTAACAACAGAGATGATTAATGTTGAAAATCCAAAT 4973
Oy 4723 CCTACCTCAATGAACAAATGCGGGTGTGACGCCGATGCTAAATGCAACGAGAGAC 4782
Db 4974 CCTACTTCAAGAAAATTAATGATGATGATGATGATGATGATGATGATGATGAT 5033
Oy 4783 AGCGGCTTAACGGAATAAATCATAGCGGCTGATTAAGCCGCGATCTATCCATC 4842
Db 5034 TCAGGTACCAAGGAAACAAATCATGATGATGATGATGATGATGATGATGATGAT 5093
Oy 4843 TTGACGAGGATTTTGTCCAGCTCTAATTTCTGGGACCTCTCTCTGCTGATCTCT 4902
Db 5094 TTGATGATTTTCTGAGATTTCTCTCAACTCTTAAGAAATTCATTTTATTAATAC 5153
Oy 4903 ATGCTGATCTGTCAGCTTCAATTAATAGATGATGAT 4939
Db 5154 ATGTTAATATATACAGTTTCAATTTAAAAATGATAGG 5190

RESULT 10
AF062348 5243 bp DNA linear INV 09-FEB-2001
LOCUS AF062348
DEFINITION plasmidium falciparum strain HNI merozoite surface protein 1
precursor (msp1) gene, complete cds.

ACCESSION AF062348
VERSION AF062348.1 GI:3859842
KEYWORDS
SOURCE plasmidium falciparum.
ORGANISM Plasmidium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmidium.
REFERENCE 1 (bases 1 to 5243)
AUTHORS Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.
TITLE Sequence analysis of the MSP 1 gene of Plasmidium falciparum from Hainan, China
JOURNAL Chung Kuo Chi Sheng Chung Hsueh Yu Chi Sheng Chung Ping Tsa Chih 17
REFERENCE (5), 294-297 (1999)
AUTHORS Jiang, G., Liu, R., Daubenberger, C.A. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1998) Molecular Immunology, Swiss Tropical
Institute, Socinstrasse 57, Basel CH-4002, Switzerland
FEATURES
source
1..5243
/organism="Plasmidium falciparum"
/strain="HNI"
/db_xref="taxon:5833"
/chromosome="9"
/country="China"
/gene="msp1"
/gene="msp1"
/product="merozoite surface protein 1 precursor"
66..5150
/note="binds spectrin"
/codon_start=1
/product="merozoite surface protein 1 precursor"
/protein_id="AAC72884.1"
/db_xref="GI:3859843"
/translation="MKIIFLCSEFLPIINTQCTHSEYQELVKLBALDEAVLTGYS
LFOEKVNLNGSTGATVTTSTPGSGSVTSVSGGSAVASVSGSGSVSGSGNS
RRTPNSSSDSAKSYADLRHONYLTFLTKELYPDLPTLHMLTLCNHGIFY
LIDYEELNELLYLNFTYDILRLAKINDACANSQCPFNKIRANLDELTKTYQY
RKPLDNTKDNVGRMEDYTKKRTITANIINELBSKRTIDONKADNEBGRKTIQAO
YNLPIYNQLODEAHNLISYLEKRLDTLTKLNENIKLELIDDKITDLEKLTGSKPNP
LPEKKKEVEGHEKIKELIAKTIFENIDSLTDLELEYLRKRNKVDYPKSQDPT
KSVQIDPKVYNGIYVPLPLDINSLAADNDKSYGGLAMPDREKINENKIIDNKE
RKIFINIKKOIDLEKKNINTEKONKLEKDYKSKDYELLEKYEKMFNNPFR
DYVDKISASRYTVAVKORVNNKSSSNYSVYNOKIKKALSYLEDYSLRKGISEKDF
NHYTLTGLDEADIKKLEETKSEENKILEKRFKGLTHSANALESVDIYVLOQKYL
LIRKLEIDRLTEFLKNAQLODSIHVPITKPKNPEYIYLAKKEVDLKEIPRY
KDLKKEQAVLSITOPLVASSETTEGDSHTLSOGETELETGATTTVITIL
PTQPSPEKVEVYENSIEKSSNDNSQALTKTVYLKLEDELTFSYICHKILVNS
MDOKLLEYNLTPEKNIKSCDPLDLIFNTONNIPAMYSLYDSMNDLOHLPFELV
KEMYLHKLKEENHKLLEBQKQITGTSSTSPGATVTTAQSATHSNOOSNA
SSTNPQNAVSSGPAVESHDPVLASISNDLKGIVSLNLDNKRKVPPLITST
ESTMEKFEYENILKNNDYFENDIKQFVKSRYITGLTFOKALNDEIKKILQILSF
DLNRYKLKIDRLFNKKELEQDKMOIKKILKLEQLESKLSNSLNPNHLYONFSVE
MKKEAEIETENTLENTKILKHYKGLVKNESSPKLTLSVSIOTEDNYANLEK
FRVSKIDGKIDNHLHGKRLSFLSGIHLITELKVKIKKKYTKSGENKKYN
EALSKYENFDEPAKVTVYTPQDVPYPSPLSVYSSGSGTRETQTPSGSLTEL
QOVQSOVNYDEEDSLVPLFEGSEDNDELDOVYGEALSVTMDNLSCFENEIDY
IYLPPLAGVNSLAKQLEKNIITPNLNDLNSRLKRLKFLDVLBSLMQFHISS
NEYITIEDSFKLINSEOKNTLKSRYIKESYENDIKRAQEBISYEVKALXKDLDS
IKVYIKEREKPESSPTPTPSPATDOKKLESLPLFJNETIETLYNNLVNKIDYLLN
LKAKINDCNVPRDEAHVKTISDLKIDIXIDLPKNTNDEATIKKILINDTKSDMLG
KLUSTGLVONFPNTIISKLEGRQDMANISOBQCVKQCCPENGSCRRHIDREDC
LINTKQEDKCVENPNPCNNNGCDADACTEEDSDSGNKITTCCTKPDSTPLFD
GIFCSSSFFLEISPLILMLLYSEPI"

BASE COUNT 2303 a 677 c 686 g 1577 t
ORIGIN

Query Match 26.9%; Score 1331.2; DB 3; Length 5243;
Best Local Similarity 56.9%; Pred. No. 1.2e-281;
Matches 2919; Conservative 0; Mismatches 1978; Indels 231; Gaps 17;

OY	10	ATGAAATCATTTTCTCCCTCCCTGATATTCGTTTTATCATCAATATCCAGGCGCTG	69
Db	66	ATGAAGATCATATTTCTTTATGATGTCATTTCTTTTTATATPATAATACAAATGTSTA	125
OY	70	ACCACGAATCTATTCAGAGAGCTGGTTAAGAAACTGGAACTTTGGAGAGATCCGCTCT	129
Db	126	ACACATGAAAGTTTCAAGAACTGTCCAAAACCTAGAAAGCTTTAGAAAGATSCAATATTG	185
OY	130	ACCGGATACACCCCTGTTCCAGAAAGAGAGATGTCTGAATGAAGGACGAGTGGCACG	189
Db	186	ACAGGTTATATGTTATTTCAAAAGGAAAAATGTTTAAATGAAGAAACAAGTGGACACA	245
OY	190	GCCGTTACAACACACACCCCGGTCTMAAGGGTCTGTGGTAGGGGTGGCTCGGGTGG	249
Db	246	GCTGTTAACATATATACACCTGGTTCAGGTGTTCCAGTTACTCAGGTGGTTCAGGTGT	305
OY	250	TCTGTGGCCTCTGGGGGTTCGCTCGCTCCGCGGAGCGGTGGCATCAGGTGGCTCAGTG	309
Db	306	TCAGTTGCTTCAGTTGCTTC-----AGGTGGTTCCAGTGGCTCACTT	347
OY	310	GCAMCGCGCGTTCGCGGAACAGTCAGAAACCAATCCATCTGACACACTTAGCGATTCC	369
Db	348	GCTTCAGGTGGTTCAGGTAATTCAGACGCTACAAATCTTCAGATATTCAGATATTCAC	407
OY	370	GACGCGAAGTCTCTACGCGCGACCTCAAGCACACGAGTGAAGAACTATCCTCCATCTCAG	429
Db	408	GAITGTAATATTTACGCTGATTTAAACATATAGCTTCAAAATTAATCTTGTCCTATTAA	467
OY	430	GAGCTGAAGTACCACAGTGTGTCGACCTCATAATCATATGCTGACACTGTGTATTAAC	489
Db	468	GAAGTCAAAATATCCGGAACCTTTGATTTAACCAATCATATGTTAACTTGTGTATTAAT	527
OY	480	ATTCAATGCTTCAAAATTCGTGATTGACGGTTACGAAGAGATCAATGAACTCCTGTACAG	549
Db	528	ATTCAATGCTTCAAAATTTTAATGATGATTAAGAAATTAATCAATTAATTAATATAA	587
OY	550	TTGAATTTCTCTCGATCGATGATGAAGGGCCAACTGAATAGCGTTGGCCCAATGACAT	609
Db	588	TTAACTTTTATATATGATTTATTAAGAGCCAAATTAATATGATGATGCTCCAAATGTTAT	647
OY	610	TGTCAAATTCATTCATTTTGAAGATCAGAGCCACAGAGTGTGAGCTATTGAAGAAGTTG	669
Db	648	TGTCAAAATCCTTTCATCTTAAATTCGTGCAAAATGATTTAGAGCTACTTAAAAAATTT	707
OY	670	GTCCTCGATATGCGCAAGCCTCTCGACACATCAGAGACATGTGGGAAAGTGAAGAT	729
Db	708	GTCGTCGATATGAAGAAACCATTAAGCAATTAATTAAGATTAATGTAGAAAAATGAGAAAT	767
OY	730	TATATTAAGAAATTAAGAAACCATCGAGAACATTAACGAGCTGATGCAAAATCCMAA	789
Db	768	TACATTAAGAAATTAAGAAACCATTAAGCAAAATTAATTAAGATTAATTAAGAAAGTATAG	827
OY	790	AAGACCATTAAGCAAAATTAAGAAATGCAACACAGAGAGAAAGAAAGAAAGTTGTACAG	849
Db	828	AAAACAAATTAATTAAGAAATTAAGAAATGCGATTAATTAAGAAAGAAAGAAAGAAATTAATTA	887
OY	850	GCCGAGTACGAGCTGTCCATCTATTAACAAACAGCTTGAAGAAAGCCCATTAATCTATCAGC	909
Db	888	GCTCATATTAATCTTTTATTATTACAAATTAACATTAACAAAGAGACATATTAATTAATAGC	947
OY	910	GTAACGGAAGAGCGCATTAAGACACCCTCAACAACAATTAAGAAATTAATCAAGAACTGCTGCAC	969
Db	948	GTTTATTAAGAAAGCTATTAAGCACTTTAAAAAAAATGAAGAAACCTTAAGAAATTAATCTTGA	1007
OY	970	AAGATTATTAAGAAATTAAGA--ATCTCCGCGCAGCCACCTGTGGGAACACCCCTTAACAG	1026
Db	1008	GAATATGATTAATTAATTAACAGATGCCGAAAACTCACTACTGTGAAGTTAAACCAATCTT	1066
OY	1027	CTGC---TGGAACAAGAACAGAGATAGAGAGACAGAGAAAGAGATCAAGAGATCTGCC	1083
Db	1068	CTCCCTGAGATTAAGAAAAAAGAGTGTAGAGGACACGAGAAAGAAATTAAGAAATTAATTTGCC	1127
OY	1084	AAAACATTAAGTTCAACATAGATTTCTCTTACTGATCCCTTGAGCTGGAATGATAC	1143

Db	1128	AAACGTATTAATTTAATGACATTTGATAGTTATTTACTGATCCACTGATTTGAATATATAT	1187
Qy	1144	TTGAGAGAGAAATTAAGATTTATGACATCTCGCAAAAGTCG-----GACAAAG	1194
Db	1188	TTAAGAGAAAAAATAAATAAGTTGATGTATACCCCTAAATTCAGAAATCTACGAAATCT	1247
Qy	1195	GAATCAACCGAACTATATGATATATCCAAATGGTGAGCTACCTCTGCTTTATTAACGAT	1254
Db	1248	GTTCAAATACCAAAAGTTCCCTTATATCCAAATGGATATGTATATCCCTTTACCACTACGTAT	1307
Qy	1255	ATCAACA-----CGCTCTCAACGAGCTCAATAGCTTGCGATGGATTTAAACCC	1305
Db	1308	ATTCATATATCATTAGCTGCGAGATTAATGATATAAATTCATATGATGTATTAATGATCCCT	1367
Qy	1306	TTTCGATTTATACGAAAGAACCCCTTAAGATATCTACACGACATGAGAGAAAGATTT	1365
Db	1368	GATACCTAAAGAAAAAATTAATGAAAAAATTTATTCAGATTAATAGAAAAAATAATATCT	1427
Qy	1366	ATCAACGAAATCAAGAGAGATCAAAATTTGAGAAAGAAAAATTTGAGAGTACAGAA	1425
Db	1428	ATTATATATACATTAATAAACAATTTGATTTAGAAAGAAAAAACAATTAATACACAAAGAA	1487
Qy	1426	AGTTACGAGAGCCGACGACAAAGCTCTTAACGATATCACTAAGAGATGAAAGCTGTG	1485
Db	1488	CAAAATATAAAAAATTCGAGAGATTTATGAAAGTCAAAAAAGAAAGATTTAGAAATTA	1547
Qy	1486	AACGATCTATGATTCGCAAAATTCACCAATTAATCATGACCTGACCACTGTGAGAAATG	1545
Db	1548	GAAAATTTTATGAAATGAAATTTTAATATATTTTATGCAAAAGATGCTGTATATTAATA	1607
Qy	1546	ATGGGAAAAACGATCTCTTCAAAAGTGGAGAACTGACACACATTAATCTTTTGCAATCC	1605
Db	1608	TTCACTGCAAGATATACATATTAATGTTGAAAAACAAGATTAATATTAATTAATTTTCA	1667
Qy	1606	TATGAGATTTTGAAGCATTAATCTTGAGAGCTTACCAAGCTCTTAAGTATATGAGAGAC	1665
Db	1668	TCTATATATTTCTGATATTAATGTTCAAAAATTTAAAAAGGCTCTTCAATCTTGAAAT	1727
Qy	1666	TATCTCTCGCGAATCTGTTGGAGAAAGAACTAAATTAATTAACAAGATCTCTAATG	1725
Db	1728	TATCTTTTAAAGAAAGAAATTTCTGAAAAAGATTTTAATCATATTAATTAATCTTGAAACT	1787
Qy	1726	AAGATCGAAAAACGATCGAGACGCTTGTGAGACATTTAAGAGATTAAGAAACAGTTG	1785
Db	1788	GGCCCGAAGCTGATATTAATAAATAATTAACGAGAAATTAAGAGTATGAAAAACAAAT	1847
Qy	1786	TTTGAGAGAAAGATTAACAAAAAGCAAAATTAACCGAGATGAGAAAGATCTGTGAGCTCC	1845
Db	1848	CTAGAAAAAATTTTAAAGGCTAACCAATTCAGCAAAAG---CTTCTTGAAAGTATCT	1904
Qy	1846	GATATTTTAAAGTCAAGAGCAAGAGCTCTCCCTATACACAAATTTGATGAACTCAAG	1905
Db	1905	GATATTTGTAATTAATTCAGATTCAAAAAGTTTATTAATTAATAAATAATGAAAGACTTAAG	1964
Qy	1906	AAGACTCACTCATTTCTGAAGAAACGTGAGATTAAACATTAATATCATGTGCGGATATGT	1965
Db	1965	AAGATAGATTAATTTTAAAAAATGCAACACTAAAGATAGATTAATCATGATCAACAAAT	2024
Qy	1966	TATTAAGCAGGAAATTAAGCAGAAACCATACTACTCATGCTGCTCAAGAAAGATATGAC	2025
Db	2025	TATTAATACCAAAATTAACCAAGAACCTATTAATTAATGTATTAATAAATAAGAGATGAT	2084
Qy	2026	AAACGAAAGTTCATGAGCCCAAGTCTGAGAGCTGATTAACAGAGAGAGAAAGAACTT	2085
Db	2085	AAATTTAAAAAGATTTATATCCAAAAAGTAAAAAGACATGTTAAA--GAAAGACAAAGCTGT	2142
Qy	2086	AAAACTGAGAGCAGTCAAGTAACTCCGAGCTTCCACGAAAGAGAGATTAACGGACAG	2145
Db	2143	TATCAATATTAACACAACTTTAGTTGACACAAAGCAAACTCAAGAT--GGGGGTAC	2201
Qy	2146	GCTACCAACCAAGCCGAGCAACAGAGCCGGTTCAGCTCTGGAAGGGATATAGCTGCAAGCT	2205

Db 2202 TCACACACATATATCCCAATCAGGAGAAACAGAAAGAAAGAAACAGTAGACAC 2261
QY 2206 CAGGACACAGAGACAGACAGGACCTCCAGTGCACAGTCCCGCTTCAGAGCTTAA 2265
Db 2262 ACAACAAAGGTAACATATACATTTACCACCAACACACATACACCCACCAAAACATTA 2321
QY 2266 GCTCAAGTGCCTACACACAGGCTCCTGATATACAGACGAGCAATGTGCAACCTG 2325
Db 2322 GTTGTGTAATTCATATGAAATTAAGAGTAATGACAAATTCACAGCCCTTGCAAAAACA 2381
QY 2336 GACTACCTTGAGAACCTCTATGATTCCTGATATACCTCATCTGACCAAAATATTC 2385
Db 2382 GTTATCTAAGAAATTTAGATGAATTTTAACTAATCATATATATGATCAATTAATAT 2441
QY 2386 CTCGCTCTCAGCAGCATGTGACAGAGAAGATTTCTTAAACAGTACAAAGTAAACAGAA 2445
Db 2442 TTAGATTAACACTAGATGAGACCAAAATTTATGAGATTAATTAATCTTACTCCAGAA 2501
QY 2446 GAGGAGATTAACCTGCTCTGATGACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 2505
Db 2502 GAAAAAATGATTAATTAATTCATGATTCATGATTTATTTATTAATTAATTAATTAAT 2561
QY 2506 ATTCCTGTTATGATATGATGTTGATGATGCTCAACAAATCTCTCTCAACCTGTTGATG 2565
Db 2562 ATACCTGCTATGATATGATATGATATGATATGATATGATATGATATGATATGATAT 2621
QY 2566 GAGATATATGAGAGATATGCTGCAACCTGTAATTAACCTCAAGACAGACAGAAAT 2625
Db 2622 GATATATATCAAGAAATGATATTTATTTATTAATTAACCTCAAGAGAAATATCATTC 2681
QY 2626 AAGAACTTCTGAGG----- 2641
Db 2682 AAAAAATTTATGAGAGACCAAAACAAATTAACCTGAGACATCATATCCAGTCTGGA 2741
QY 2642 -----AAGCTAAGAGGCTCTCAGCTCTGTTTAAACCTCTCTCTCAGCTCATCAAC 2696
Db 2742 AATACAAACCGTAATATGCTCAATCCGCAACTCAGATTAATCCCAAAACCAACATCA 2801
QY 2697 ACTGCTCTCAGACCTCAAGACA----- 2730
Db 2802 AATGATCTCTCAGCAATACCAAAATGCTGATATCATCTGCTGCTGCTGCTGCTGCTG 2861
QY 2721 AGCGCTAAGACGACCTCTCATCGACCAACCTTAATTAACCTCAAGAACTGTTGAG 2790
Db 2862 GAAGAAATCATGATCCCTTAACATATGCTATGCTATGATAGATGTTGAAGATGTTG 2921
QY 2791 AACATCCTGCTCTCGGCAAGATTAAGAACT---CTACCAAGAACTTATGACAGAA 2847
Db 2922 AGTCTCTTAATCTTGAAATTAACCTAATCCATTAACCAATTTCTTCAACA 2981
QY 2848 TCGTCCGAGAACTCTAGAGAGATGATGAAAGACAGACACATCTATTAAGAGAGC 2907
Db 2982 GAGATGAAAAATTTATGAGATATTTTAAAAATTAATGATACCTATTTTATATGAT 3041
QY 2908 TTCCTAATCTGCTGAAATCTAAAGCCGATGATATCACTCTTAAGATGAATCTAA 2967
Db 3042 ATCAAACTTCGTAATCTAATCAAAAGTAATTAACAGGTTGAC---GAACACAA 3098
QY 2968 CGTAAGAGCTGGAAGAGACATCAATAAGCTGAAGAGACCTGCAACTGAGCTTGAC 3027
Db 3099 AAAAAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3158
QY 3028 CTGTACACAGTACAACTGAACTGAGAGACCTTCGACAAAGAGACAGCTGCGC 3087
Db 3159 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3218
QY 3088 AAGTATTAAGATGACAGATCAAGAAATGATCTCTCAAGAGACCTTGAAAGCAAACTC 3147
Db 3219 CAAGCAAAATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3278
QY 3148 AACTGATGAAACATCCGAAACACTACTGAGAACTTCTGATGTTCTTCAACAAGAA 3207
Db 3279 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3338

QY 3208 AAGAAAGCCGAGATGCGGAGACAGAGAAACATCTGAGAAACCAACCAAGATTTCTTCA 3267
Db 3339 AAGAAAGCTGAATATGCAAGAACTGAAACACATTAAGAAACACAAATATTAATTA 3398
QY 3268 CACTCAAGAGGCTGCTGATATTAATGAGCGAGTCTCTCTCAAGACCTGCTCC 3327
Db 3339 CATTTAAAGGACTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3458
QY 3328 GAGGAGACATTCAGACCGAGAGATTAACGCGACCTCGAGAACTTCAAGTCTGCT 3387
Db 3459 GAAGTATCAATTAACAAAGAGATTAATGATCAATTTAGAAAAATTTAGAGTATTA 3518
QY 3388 AAGCTGGAAGGCACTGAGAGACAACTGAACTCGGAGAGAAAGAAAGAACTGAGCT 3447
Db 3519 AAAAAATGAGAACTCAATGATATTAATTAATTAATTAATTAATTAATTAATTAAT 3578
QY 3448 TCTAGGAGACATGACCTGATCCGAGCTCAAGAGAGTCAATTAAGACAAAGATAC 3507
Db 3579 TCAATGATTAATCATCATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 3638
QY 3508 ACCGCAATTAAGCCAGAGAGATTAATACAGCTGAATTAACGACCTGGAATCTTAAG 3567
Db 3639 ACAGTAAATTCCTCAAGGAAATTAATTAAGAAAGTTAACGAAGCTTAAAACTTACGA 3698
QY 3568 AAGTCTGCGCTGAGAGAAAGATGCTGC----- 3596
Db 3699 AATTTTCTCCAGAAACAAAGTTTCAACAGTTGTAACCTCCACCTCAACAGATTAAT 3758
QY 3597 CACTGTGCTGCTGATGCTGCTCCGACACTGAGAGAGTCACTCACTAAGAA----- 3650
Db 3759 CCATCTCATTAATCTGTAAGGTAAGTGTAGTTCAGATTCACAAAGAAACACAA 3818
QY 3651 --GCTGATCTACTCATGCTGAGAGCCAGTCAATTAATTAACCAATCTCAAGAGCT 3708
Db 3819 ATACCACTTCAAGCTCTTATTAACAGAAATTAACAAAGTAATGATCAATCAAAATTA 3878
QY 3709 GAGGATGAGGCTGAGAGCTCATCTGCTATCTGCGGAGAGAGAGAGAGCTAC 3768
Db 3879 GAGCAAGAGATGATTCCTTAATGTTATTAATTAATTAATTAATTAATTAATTAAT 3938
QY 3769 GATGACCTGCGGAGAGGCTGACCGTGAGGCTGCTCTCCGATTAATTAATTA 3828
Db 3939 GAATATTTAGATCAAGTATGATTAATTAATTAATTAATTAATTAATTAATTAAT 3995
QY 3829 CTGCTCAAAATCGAAGAGATTAAGAGCTGCTTAATTAATTAATTAATTAATTAAT 3888
Db 3996 CTCTGAGATTTGAATGAATGATGTTATTAATTAATTAATTAATTAATTAATTA 4055
QY 3889 AGGCTCTCAAGAAACAGCTGAGAAATTAAGGATGATCAATTAATTAATTAATTAAT 3948
Db 4036 AGAAGCTTAATAAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4115
QY 3949 ATTCTGAACGCGCTTAATTAAGAGAAATTTCAAGACGCTGAGAGAGCTTG 4008
Db 4116 ATCTTAATTAATCAAGCTTAAGAAAGAAATTTCTTAAGATGATTAATTAATTAAT 4175
QY 4009 ATTCCTTAATTAAGACCTGACCTCTTAATTAATTAATTAATTAATTAATTAATTA 4068
Db 4176 ATGCAATTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4235
QY 4069 AATAAGAGAGAGAGATTAATTTCTGCTAGTTAATTAATTAATTAATTAATTAATTA 4128
Db 4236 AATTAAGAAACAAACCACTTTAAAGATTAATTAATTAATTAATTAATTAATTAAT 4295
QY 4129 ACCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4188
Db 4296 AATGATTAATTAATTTGACAGAGAGATTAATTAATTAATTAATTAATTAATTAAT 4355
QY 4189 AAGTCTGACCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4230
Db 4356 AAGGATGATTTAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4415

OY	310	GCACGCGCGGTTCCGGAAACAGTCCAGAACCAATCCATCTTGACAACTGTAAGCATTC	369
Db	283	GCCTTCAGGTGGTTCAGTAATTCAGACGCTACAAATCTTCAGATAATTCACGATTC	342
OY	370	GACCCAAAGTCCTACGGCGACTCAAGCAACCGGTGAGAAACTATCTCTCTCATTCAG	429
Db	343	GATCTAAATCTTACGCTGATTTTAAACATAGAGTTCCAAAATTACTTGTCCTATTTAA	402
OY	430	GAGCTGAAGTACCCAGAGTTGTTCGACCTCACTAATCATATGCTGACACTGTGTATAC	489
Db	403	GAAGTCAAAATATCCCGAAGCTCTTGATTTTAAACAAATCATATGTAACTTTGTGTATAT	462
OY	490	ATTATGCGCTTCAAAATATCTGATTTGACCGGTTACGAAGATCAATGAACCTCTGTACAG	549
Db	463	ATTCATGGTTTTCAAATTTTAAATTTGATGATTTGAAGAAATTAAGAAATTTATATATAA	522
OY	550	TTGAATTTTACTTCGACTGCTGAAGGGCAAACTCAATGAGCTGTGGCCAAATGACTAT	609
Db	523	TTTAACTTTTATATGTATTTATTATAGAGCCAAATTAATGATGCAATGTGCCAATATGTTT	582
OY	610	TGTCAAAATTCATTCATTTTGAAGATCAGAGCCAAACGAGTGTGAGCTATTTGAAGACTTG	669
Db	583	TGTCAAAATTCATTCATTTTGAAGATCAGAGCCAAACGAGTGTGAGCTATTTGAAGACTTG	642
OY	670	GCTCTCGGATATGCGCAAGCCTCTGCACAAATCAGAGCAATGTGGGAAGATGAGAT	729
Db	643	GCTCTCGGATATGCGCAAGCCTCTGCACAAATCAGAGCAATGTGGGAAGATGAGAT	702
OY	730	TATATTTAAAGAAATTAAGAACCAATTCGAAACATTAAGAGCTGATGCGAAGATCCAA	789
Db	703	TATATTTAAAGAAATTAAGAACCAATTCGAAACATTAAGAGCTGATGCGAAGATCCAA	762
OY	790	AAGACATAGACAAAATTAAGAAATGCAACCCAAAGAGAAAGAAAGAAAGATGTACCAG	849
Db	763	AAACAAATGTGATCAAAATTAAGAAATGCAAGATTAAGAAAGAAAGAAAGATGTACCAG	822
OY	850	GCCCAATGACGCTGCTCATCTATTAACAAACAGCTTGAAGAAAGCCATTAAGCTATACG	909
Db	823	GCTCAATATATCTTTTATTTATTCATTAACAAATTAAGAAAGCAATTAATTAATTAAGC	882
OY	910	GTAATGAGAAAGCGCATAGACACCCCTCAAGAAATGAAGAAATATCAAGAAATGCTGAC	969
Db	883	GTTTATGAAAAACGATATGACACTTTTAAAAAAAATGAAACATTAAGAAATTAAGCTTAA	942
OY	970	AAGATTAATGAATTTAGA--ATCCTCGCCAGCCAACTGTGGAAACACCCCTAACAG	1026
Db	943	GATATGATTAATTAATTAAGAAATGAGATGCCGAAAACTCACTAGTAAGATTAACCAATCT	1002
OY	1027	CTGC---TGGACAAGAACGAAGATATGAGAGCACGAGAAAGATCAAGAGATGCC	1083
Db	1003	CTCCCTGAGATTAAGAAAAAGAAAGTCCGAGGACGAGAAAGAAATTAAGAAATTTGCC	1062
OY	1084	AAAACATTAAGTCAACATAGATATCTCTTACTAGATCCCTTGAGCTGAGATGATAC	1143
Db	1063	AAAACATTAAGTCAACATAGATATCTCTTACTAGATCCCTTGAGATGATGATGATGAT	1122
OY	1144	TTGAGAGAAAGATTAAGATATAGACATCTCCGCAAAAGTGA-----GACAAG	1194
Db	1123	TTTAAAGAAAAAATTAAGAAAGTGTGATGTATACACTTAATTCACAAAGATCTTACGAATCT	1182
OY	1195	GAATCAACCGAACCTTAATGAATATCCCAATGTGTGACGTACCCCTGTCTTATTAAGAT	1254
Db	1183	GTTCAAAATTAAGTCTTCTTATCCAAATGTGATGTATTCCTTTTACCACTGAT	1242
OY	1255	ATCAACA-----CGCTCAACGAGCTCAATAGCTTGGGTGACTTGATTTAACCC	1305
Db	1243	ATTCATATATTCATAGCTGACAGATATGATTAATAATTCATATGATGATTTATATGATCT	1302
OY	1306	TTTCAATTAAGAAAGAACCTCTGAAGATATATACAGACAGATGAGAAAGAAAGTTT	1365
Db	1303	GATACATAAGAAAAAATTAATGAAAAATTTATACGATTAATTAAGAAAGAAAAATTAATTC	1362
OY	1366	ATCAACGAATCAAGAGAAATCAAAATTTGAGAAAGAAATTTGAGAGTACAAAGAA	1425

Db	1363	ATTATAACATTA	AAAAAAAAA	CAAAATTTGATTTAGAGAAAAA	AAAAA	CAATTAACTACACAAA	GAA	1422
Qy	1426	AGTTACGAGAC	CCGACGACAA	AGCTCTAAACGATACATCA	TAAAGATATGAA	AAAAAGCTGCTG	1485	
Db	1423	CAAAATTA	AAAAATTA	ACTTGTAAGATTTATGAAAGTCA	AAAAA	AGGATTTATGMAAATTA	ACTT	1482
Qy	1486	AACGAGATCTA	TGATGTTCCA	AAATTTCAACATTAACATC	GCCTGACCA	ACTTGAGAAATG	1545	
Db	1483	GAAAAATTTT	TATGAAATGA	ATTTAATTAATTTTGACAAAGATG	CTGTAGATTA	AAATA	15442	
Qy	1546	ATGGGAAAA	CGGTACTCTT	CAAGAATGGA	AAACGTGACACACCATTA	TACCTTGCAATCC	1605	
Db	1543	TTCACTGCA	MAATVTPACAT	TAATTAATGTTGAAAAA	CAAAAAATTAATTAATTAATTTTCA	ATCC	16020	
Qy	1606	TATGAGATTT	CTAAGCTTAATCTT	GAAAGCTCACC	AAAGCTCTTAAGATATG	GAGAGAC	16655	
Db	1603	TCATAATTA	TTCTGTATATTA	TATATGTTCAAAATTTAAAAA	AGCTCTTTCATATCTT	GAGAT	16632	
Qy	1666	TATTCCTCG	CGGAACATTTGTT	GAGAAAAA	CTAAATATTAACAGATCT	CTAATAGT	17255	
Db	1663	TATTCCTTA	AGAAAGGAATTTCTG	AAAAAGATTTTAATCATTAATTA	TACTTGAAATCT	17222		
Qy	1726	AAGATCCAAA	ACGATCGAGCGCTT	GTGAGACATTTAAGAAGATGA	AGACAGTTG	17855		
Db	1723	GGCCTCGA	AGCTGTATTA	AAAAAATTAACGAAAGAAATTA	AGATGAGAAAA	CAAAATTT	17822	
Qy	1786	TTTGGAGA	GAGATTTACAAA	AGACGAAATTAACGATGAG	AGAAATCCTGAGAGCTG	18455		
Db	1783	TTAGAAAAA	ATTTTAAAGAC	CTAACACATTCAGCAATG	---	CTTCCTTGA	AGTATCT	18399
Qy	1846	GATATTTG	TTAAAGTCAAGTG	CGAGAGTGCTCTCATG	ACACAGATTGATGA	CTAAG	19055	
Db	1840	GATATTTGA	AAAAATTAACAGTA	CAAAAAATTTTATTAATTA	AAAAAATGAA	AGACTTA	GAA	18999
Qy	1906	AAGACTCA	ACCTCTTGAGA	ACGTGAGTTAAACATTAATATCAT	GTGCGCAATAGT	19655		
Db	1900	AAGATAGATTA	TTATTTTAAAAA	ATGCAACATTAAGATGATTA	CTCATGTACCAATAT	TTT	19599	
Qy	1966	TATATA	CGAGAGATTAACAG	AGACCATATACCTCATCTCA	CTCAAGAGATG	ATAGC	20255	
Db	1960	TATATA	CCACCAAAATTAAC	CAAGAACCATATTTATTAATGTATTA	AAAAAAGACT	ATAT	20199	
Qy	2026	AAACTGAA	AGTGTTCATGCC	CAAGTGGAGGCTGATCA	CGAAGAGAC	AGAACATTT	20855	
Db	2020	AAATTA	AAAAATTTATATCC	AAAAAGTAAAGACATGTTAA	-GAAAGCA	MAAGCTGTCT	20777	
Qy	2086	AAAACTGA	AGACACGTAGATTA	CTCGAGCCTCCACAGA	AGAGATTA	ACGAGAC	21455	
Db	2078	TATCAAGTA	TTATACAA	CACTTTAGTTGAC	CAAGCAAGCAAACTGA	AGAT-GGGG	GTAC	21366
Qy	2146	GCTACAC	CAAGCCCGGAC	CAAGCGCGGTGCTCTG	AGGGGATATAG	CGTCACT	22055	
Db	2137	TCACAC	ACACATTTATCC	CAATGAGGAACGAAGAAACGA	AAACAGT	AGGACAC	21966	
Qy	2206	CAAGAC	CAAGACGAGAC	AGGACAGCCTCCAGTCC	AGTCCCGTTC	CAAGAGCTTAA	22655	
Db	2197	ACAACA	ACGGTAACAA	ATTAACATTTACCA	CAACACACATCA	CAACAAAAAGAA	AGATTA	22566
Qy	2266	GCTCAAGT	GCCTACACCA	CAAGCTCCTGTG	AAATACAGACCGAG	ATGTACGCAATG	23255	
Db	2257	GTTGTG	AAAAATTTCAAT	AGAAACATTAAGATTA	AGCAATTTCA	CAAGCCTTG	CAAAAAA	23166
Qy	2326	GACTATCC	TTGGAAGCTCT	TGATGATTCGTAATAC	ATACATCTCAT	CGCCACAAATATATC	23855	
Db	2317	GTTATCTTA	AGAAATTAATGAT	TGAATTTTAACTTAATTAATTA	TATATATGATATATATATAT	TTT	23766	
Qy	2386	CTCGTCTCT	CACACCTATATGA	ACGAGAGATTTCTTAACAGATTA	ACAGATTAAC	CAAGAA	24455	
Db	2377	TTAGATCA	AACTCTATAT	TATGACCAAAAAATTTATG	AGATATATATCTT	ACTTCAGAA	24366	
Qy	2446	GAGGAG	ATAAACTGCTCTT	GTGTGATCCATG	GCCTGCTGTCAATAT	TCAGAACAC	25055	

[illegible]

QY	3508	ACGGCATTACCCACACCGAGAAATAATACGACGTAAATATACGCACTGGAATCTTACAG	3567
Db	3574	ACAGGTAATTCCTCCAACTGAAATAATATAGAAAGTTATACGAAGCTTTAAATCTTACGA	3633
QY	3568	AAGTTCGCGCTGGAAGAACAGATGTCGC-----	3596
Db	3634	AATTTTCTCCAGAAAGCAAAAGTTACAAACAGTTGTAACTCCACCTCAACCAATGTAACT	3693
QY	3597	CACGTGCTGTCTAAATCTGGCTCCGACACACATCGAGCACTGCAACCTAAGA-----	3650
Db	3694	CCATCTCCATTAATCTGTAAAGGGTAAGGTGATGTCAGGATCCACAAAGAAACACAA	3753
QY	3651	--GCCGTATCTACTCAATGTCGGAGCCGAGTCCAAATACATTAACACATCTGAGACGTC	3708
Db	3754	ATACCACTTACGCTCTTTATTTACGAGATTTCAACAAGTAGTACATACACAAATATAT	3813
QY	3709	GACGATAGGTGATGATGACGTCATCATTTGTCCTATCTTCGGCGAGACGAGAGCATAC	3768
Db	3814	GACGAGAAAGATGTATCTTACGTTAGTTGTATTCACCATTTTGGAGAAATCCGAAATATAC	3873
QY	3769	GATGACCTCGGCCGAGGTGGTCACCGGAGACGCTGCATCTCCGTGATATACAT	3828
Db	3874	GAATTTTAGTCAAGTAGTACTGGAGACCATATCTGTCAATG---GATAAATTC	3930
QY	3829	CTGTCCAAATCGAAGACGAATACGAGTGCCTATCTGAACCTCTGGCAGCCCTAT	3888
Db	3931	CTCTCAGATTTGAAATGAAATGATATGATATATTTTAAACCTTACTGAGATATAT	3990
QY	3889	AGGTCTCCAGAAACAGCTGGAGAAATACGTATACCTCTCAATGTCACTGAGAGAC	3948
Db	3991	AGAACTTAAAAAACAATTTGAAAAAACATTATTTACATTTAAATTTAATTTGAAGAT	4050
QY	3949	ATTCTGAACACCCCTTTAATTAAGAGAAATTTCAACAGCTCTTGGAGACACTTG	4008
Db	4051	ATCTTAATTCACGCTCTTAAGAAACGAAATATTTCTTGAGATGATTAAGATCTGATTTA	4110
QY	4009	ATTCCTTAATAAGCTGACCTCTCTACTACTGTTGTCAAGACCATACAAAGTTCTC	4068
Db	4111	ATGCAATTTAAACATATATCTCCAAATGAATATATTATTTGAAGATTCATTAAATATG	4170
QY	4069	AATAAGAGAGAGAGATTAATTTCTGTATACACTATATACAGATCCATCGAC	4128
Db	4171	AATTCAGACAAAAAACAACACCTTTTAAAGTTACAAATATATTAAGAAATCACTAGAA	4230
QY	4129	ACCGATATCAATTCCTTAATGATGTCTGGGTATATACAGATCCTGAGCGAAATAC	4188
Db	4231	AATGATATTAATTTGGACAGGAAGTATAGTATTATGAAAAAGTTTACCGAAATAT	4290
QY	4189	AAGTGTGACCTGTGCTCATTTAAAAAAGTATATCAAGATAAG-----	4230
Db	4291	AAGGATGATTTAGATCAATTTAAAAAAGTTATCAAGAAAGAAAAAGAGTCCCATCATCA	4350
QY	4231	-----CAAGCGAGATGAATAATAT	4251
Db	4351	CCACCAACAACACCTCCGTACCAACGAAAAACAGACGAACAAGAAAGAAAGTAAATTC	4410
QY	4252	CTGCGCTTCGGAATATACATCGAAACCCCTTACAAACACAGTAGACGACAAATTCACCTC	4311
Db	4411	CTTCATTTTAAACAAACATTTGAGACCTTATTAACAAATACCTTAGTTAATTAATTTAGCAT	4470
QY	4312	TTTCGTAATTCACCTGGAGGCGCAAGGTCTCACTACTATTACGAGAGAGCAATGTGAA	4371
Db	4471	TACTTAATTAACCTTAAGGCAAGGATTAACGATTTGTATATGTTGAAAAAGATGACACAT	4530
QY	4372	GTTAAATCAAGAGCTGAACCTACCTCAAAACAAATCAAGACAGCAAGCTGGCAGATTTCAG	4431
Db	4531	GTTAAATAACTAAACTTAGATTTTAAAGCAATTTGATGACAAAAATAGATCTTTTAAA	4590
QY	4432	AAAAATTAACATTTTCGTCGGAAATTTGCAGACCTGTACCGATTTATACACACAAATCTC	4491
Db	4591	AACACTTAACGCTTCGAAGCAATTTAAAAATTTATATAATGATGATATCGAAATAAAGATATG	4650

QY	3508	ACGGCATTACCCACACCGAGAAATAATACGACGTAAATATACGCACTGGAATCTTACAG	3567
Db	3574	ACAGGTAATTCCTCCAACTGAAATAATATAGAAAGTTATACGAAGCTTTAAATCTTACGA	3633
QY	3568	AAGTTCGCGCTGGAAGAACAGATGTCGC-----	3596
Db	3634	AATTTTCTCCAGAAAGCAAAAGTTACAAACAGTTGTAACTCCACCTCAACCAATGTAACT	3693
QY	3597	CACGTGCTGTCTAAATCTGGCTCCGACACACATCGAGCACTGCAACCTAAGA-----	3650
Db	3694	CCATCTCCATTAATCTGTAAAGGGTAAGGTGATGTCAGGATCCACAAAGAAACACAA	3753
QY	3651	--GCCGTATCTACTCAATGTCGGAGCCGAGTCCAAATACATTAACACATCTGAGACGTC	3708
Db	3754	ATACCACTTACGCTCTTTATTTACGAGATTTCAACAAGTAGTACATACACAAATATAT	3813
QY	3709	GACGATAGGTGATGATGACGTCATCATTTGTCCTATCTTCGGCGAGACGAGAGCATAC	3768
Db	3814	GACGAGAAAGATGTATCTTACGTTAGTTGTATTCACCATTTTGGAGAAATCCGAAATATAC	3873
QY	3769	GATGACCTCGGCCGAGGTGGTCACCGGAGACGCTGCATCTCCGTGATATACAT	3828
Db	3874	GAATTTTAGTCAAGTAGTACTGGAGACCATATCTGTCAATG---GATAAATTC	3930
QY	3829	CTGTCCAAATCGAAGACGAATACGAGTGCCTATCTGAACCTCTGGCAGCCCTAT	3888
Db	3931	CTCTCAGATTTGAAATGAAATGATATGATATATTTTAAACCTTACTGAGATATAT	3990
QY	3889	AGGTCTCCAGAAACAGCTGGAGAAATACGTATACCTCTCAATGTCACTGAGAGAC	3948
Db	3991	AGAACTTAAAAAACAATTTGAAAAAACATTATTTACATTTAAATTTAATTTGAAGAT	4050
QY	3949	ATTCTGAACACCCCTTTAATTAAGAGAAATTTCAACAGCTCTTGGAGACACTTG	4008
Db	4051	ATCTTAATTCACGCTCTTAAGAAACGAAATATTTCTTGAGATGATTAAGATCTGATTTA	4110
QY	4009	ATTCCTTAATAAGCTGACCTCTCTACTACTGTTGTCAAGACCATACAAAGTTCTC	4068
Db	4111	ATGCAATTTAAACATATATCTCCAAATGAATATATTATTTGAAGATTCATTAAATATG	4170
QY	4069	AATAAGAGAGAGAGATTAATTTCTGTATACACTATATACAGATCCATCGAC	4128
Db	4171	AATTCAGACAAAAAACAACACCTTTTAAAGTTACAAATATATTAAGAAATCACTAGAA	4230
QY	4129	ACCGATATCAATTCCTTAATGATGTCTGGGTATATACAGATCCTGAGCGAAATAC	4188
Db	4231	AATGATATTAATTTGGACAGGAAGTATAGTATTATGAAAAAGTTTACCGAAATAT	4290
QY	4189	AAGTGTGACCTGTGCTCATTTAAAAAAGTATATCAAGATAAG-----	4230
Db	4291	AAGGATGATTTAGATCAATTTAAAAAAGTTATCAAGAAAGAAAAAGAGTCCCATCATCA	4350
QY	4231	-----CAAGCGAGATGAATAATAT	4251
Db	4351	CCACCAACAACACCTCCGTACCAACGAAAAACAGACGAACAAGAAAGAAAGTAAATTC	4410
QY	4252	CTGCGCTTCGGAATATACATCGAAACCCCTTACAAACACAGTAGACGACAAATTCACCTC	4311
Db	4411	CTTCATTTTAAACAAACATTTGAGACCTTATTAACAAATACCTTAGTTAATTAATTTAGCAT	4470
QY	4312	TTTCGTAATTCACCTGGAGGCGCAAGGTCTCACTACTATTACGAGAGAGCAATGTGAA	4371
Db	4471	TACTTAATTAACCTTAAGGCAAGGATTAACGATTTGTATATGTTGAAAAAGATGACACAT	4530
QY	4372	GTTAAATCAAGAGCTGAACCTACCTCAAAACAAATCAAGACAGCAAGCTGGCAGATTTCAG	4431
Db	4531	GTTAAATAACTAAACTTAGATTTTAAAGCAATTTGATGACAAAAATAGATCTTTTAAA	4590
QY	4432	AAAAATTAACATTTTCGTCGGAAATTTGCAGACCTGTACCGATTTATACACACAAATCTC	4491
Db	4591	AACACTTAACGCTTCGAAGCAATTTAAAAATTTATATAATGATGATATCGAAATAAAGATATG	4650

Oy	4492	CTGACCAAGTTTCTGTCCACGTGCAATGSGTGTTCGAAAACCTCGCCAAAACAGTGTGAGC	4551
Db	4651	CTTGGCAAAATTAAGTATGATACAGGATTTAG--TTCAAAATTTCTTAATATCAATAATATCA	4707
Oy	4552	AATCGCTCGACGGCAACCTCGAGGGCATGCGTGAACATCTCCACACCAATGCGTGAAG	4611
Db	4708	AAATTAATTGAAGGAAATTCACAGATATGTGTAATAACATTTCACAACACCANTGCTTAAA	4767
Oy	4612	AAACAGTCCCCCAGATATGCGGCTGTTTCAAGGCATCTGGACGAGCGCGAAGAGTGCAG	4671
Db	4768	AAACATATCCAGAAAATTCGTGAGTGTTCAGACATTTAGATGAAAGGAAGAAATGTAAA	4827
Oy	4672	TGTCTCCGAATCTACAAACAAGAAAGAGATATAGTCCGGAGAAACCAACCTCTACGCG	4731
Db	4828	TGTTTATTAAATTATTAACAAACAAGAGGTATTAATATGTGTAATAATCAAAATCTTACTGT	4887
Oy	4732	AATGAAAACAATGGCGGGGTGTGACCGCATGCTAAATGCACCGAGGAAGACGCGCTCT	4791
Db	4888	AACGAAATTAATGTGGATGTGATGTCAGATGCGCAATGTACCGAAGAAAGATTCAGGTAGC	4947
Oy	4792	AACGGAAGAAAAATACATGGGAGTGTACTAAGCCCGACTCTATACCATCTTGACGGG	4851
Db	4948	AACGGAAAAAATACATGTGAATGTACTAAACCTGATTTCTTATACCACTTTTCATGTGT	5007
Oy	4852	ATTTTTGTCTCAGAGCTCAATATTTCTCGGGACATCTCTTCTCGCTATCATCTGATGTATC	4911
Db	5008	ATTTCTTCGAGTTCCTCTACTACTTCTTAGAGAAATATCATTTCTATTATTAATCATGTTAATA	5067
Oy	4912	CTGTACAGCTTCATCTAA	4929
Db	5068	TTATACAGTTTCATTTAA	5085

RESULT 12
 PEP190G1
 LOCUS PEP190G1 5392 bp DNA linear INV 03-JUN-2002
 DEFINITION Plasmodium falciparum p190 gene for surface antigen precursor.
 ACCESSION X05624
 VERSION X05624.1 GI:4186080
 KEYWORDS antigen; glycoprotein; p190 gene; signal peptide; surface antigen.
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1
 AUTHORS Tanabe,K., Mackay,M., Goman,M. and Scaife,J.G.
 TITLE Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum
 J. Mol. Biol. 195 (2), 273-287 (1987)
 JOURNAL 88011243
 MEDLINE 3079521
 PUBMED
 REFERENCE 2 (bases 1 to 5392)
 AUTHORS Tanabe,K.
 TITLE Direct Submission
 JOURNAL Submitted (21-JAN-1999) K. Tanabe, Osaka Institute of Technology,
 Biology, 5-16-1 Ohmiya, Asahi-ku, Osaka, 535-8585, JAPAN
 REMARK Sequence revised by author
 COMMENT On Jan 26, 1999 this sequence version replaced gi:9926.
 FEATURES location/Qualifiers
 source 1..5392

CDS

```

/organism="Mazodinium fallaxiparum"
/strain="MAD20"
/db_xref="taxon:5833"
/clone_1ib="(lambda) NM1149 genomic"
136..5241
/note="p190 precursor"
/codon_start=1
/protein_id="CAA29112.1"
/db_xref="GI:4186081"
/db_xref="SWISS-PROT: P08569"
/translation="MKITIFPLCSFEFIITNQCVTHTSEYELKKLEALADAVLTGTS
LFQKRNQVINESTGIAVTVTPGSSGVSIVSGSVASVSGSGSVASGSGS
RTNVESSDSNTKATYKADLKRAKNQVLPFIKELKAPELDLTNHMLTLK
LIDGNEIDNELLLKTFYVDLLAKRINDACANSYCOIPFNLIKIRAMELDL

```

[illegible]

QY	70	ACCCACGAATCCATCAGAGACTGGTTAGAACTGGAACTTTGGAAATGCCGTCCTT	129
Db	196	ACACATGAAATTTATCAAGAACTGTGTAACAAACCTTTAGAAATGCAATATG	255
QY	130	ACCGATACAGCCTTCTCCAGAGAGAAAGTGGTGGTGGATGAAGGAGCAGCGGACG	189
Db	256	ACAGTTTATAGTTTATTTCCAAAAGAAAAATGGTATTAAATGCAAGGACAAAGTGGAC	315
QY	190	GCCGTTTACAAACGACACACCCGGTTCTAAAGGTCGTGTGGCTAGCGGTGCTCCGTTGG	249
Db	316	GCTGTTCACAACTAGACACCTGGTTCAAGTGGTTCAGTTACTTCAGGTGGTTCAGTTGCT	375
QY	250	TCTGTGGCCTCTGGGGGTTCGCTGGCCTCCGGCCGACGCTGGCATAGGTGGCTCAGTC	309
Db	376	TCAGTTGCTTAGTTGGTTCTT-----AGTGGTTCAGGTGGCTCAGTT	417
QY	310	GCAACGCGCGGTTCGCGGAMCAGTCGAAGAACCATCATCTGTGCAACTTACGATATCC	369
Db	418	GCTTCAGGTGGTTGACGTAATTTCAAGACGTCAAAATCTTCAGATTAATTCAGATGATTC	477
QY	370	GACGCCAAGTCCTACGCCGACCTCAAGCACCAGTGAAGAACTATCTCTCACTATCAAG	429
Db	478	AATACTAAACCTTACGCTGATTTAAACATAGAGTTCAAAAATTACTGTTCACTATTTAA	537
QY	430	GAGCGTAGTACCCACATGTTGTGACCTCACTAATATATGTCGACACCTGTGTATAC	489
Db	538	GAACTCAATATATCCGGAACCTTTGGATTTAAACAATATATGTTAGCTTTATCTTAAAA	597
QY	490	ATTCAATGACCTCAATATCTGATTGACGGTTACGAMAGATCAATCAATGACCTGTACAG	549
Db	598	GTTGATGGTTTCAAAATATTATTAATGATGAGATGATACAAATTAATGAATTTATATATA	657
QY	550	TTGAATTTCTACTTGACTGCTGTAGGGCCAAACTGATAGACGTTGGGCCAATGACTAT	609
Db	658	TTAAACTTTTATATGATTTTATTAAGACCCAAATTAATATGATGATGTCCAATAGTTAT	717
QY	610	TGTCAAATTCATCAATTTGAAGTCGAGGCCAAGCAGTAGTTGGACGTATGAAGAAGTGG	669
Db	718	TGTCAAATTAACCTTTCATCTTTAAATTCGTGCANATATATTAATAGACGTACTTTAAAT	777
QY	670	GTCCTTCGATATCGCAGACCTCTCGACACATCAAGACATGTGGGAAAGATGGAAGAT	729
Db	778	GTTGTCGATATAGAAAACCATTTAGACAAATTTAAAGTATATGTAGGAAAAATGGAAAT	837
QY	730	TATATTAATAAAGATAGAGACCATGTGAGACCTTTAAGAGCTGATCGAAGATCCAA	789
Db	838	TACATTTAAAAAATTAACAAACCATAGCAAAATTAATTAATTAATTGAAGAGACTAAG	897
QY	790	AAGACCATAGACAAAATTAAGATAGCAACAGGAGAGAGAAAAAGAAAGAGTTGTACCG	849
Db	898	AAAACAAATTTGATCAAAATTAAGAAATGCGAGATATGAGAAAGAAAAAATTAATTAACA	957
QY	850	GCCAGTACGACCTGTCCATCTATTAACAACAGGCTTGAAGAGGCCATTAACCTCATACAG	909
Db	958	GCTCAATATATCTTTTATTTTACAAATAACAATTAACAAGAGACACTAATTTAATAAG	101
QY	910	GTACTGGAAGAGCGCATAGACACCTCAAGAGATGAATAATTCAAAGAACTGCTGCAG	969
Db	1018	GTTTATAGAAAAAGCTATGACACTTTAAAAAAAATTAAGAAACATAAGAAATTACTTGA	107
QY	970	AAGATTATGAATTTAGA---ATCCTCCGCGACGCACTCTGGGAACACCCCTAACAG	102
Db	1078	GATATAGTATTAATTAACACAGATGCGGAAAAACCCCACTAGTAGATTAACAAATTCCT	113
QY	1027	CTGC---TGGACAAGAACAAACATATGAGAGGACAGAGAAAGAGATCAAGAGATGCGC	108
Db	1138	CTCCCTGGCAATTAAGAAAAAGAAAGTGGAGGACAGCAAGAAAAATTAAGAAATTTGCC	119
QY	1084	AAAAACATTAAGTCAACATGATTTCTCTTTTACTGATCCCTTGAGCTGGAAGTACTAC	114
Db	1198	AAAACATTAATTAATTAACATTTGATAGTTTATTTACTGATCCTACTGATTAAGATATTTAT	125
QY	1144	TTGAGAGAGAAATTAAGCAATATAGACATCTCCGCCAAAGTGA-----GACAAAG	119

Db	1238	TTAAAGAGAAAAAATTAATAAAAGTTGAGTGAACCGCTTAATCACAAGATCTACGAAATCT	1317
Qy	1195	GAACTCAACCGGAACCTTAATGAATATCCCAATGAGTGTGACAGACCTCTGTCTTAACGAT	1254
Db	1318	GTTCAAAATACCAAAAGTTCCCTATACCAAAATGATATGTATATCCTTTACCACCTACGTAT	1377
Qy	1255	ATCAACA- - - - -CGCTTCACAGACTCAATAGCTTGGGTGACTTGATTAACCC	1305
Db	1378	ATTCTAATTCATTAGCTGCGAGATAAGATAAAATTCATATGTCGATTAATTAATGAATCTT	1437
Qy	1306	TTTCGATTATACGAAGAACCCTTAACAATATGTACACAGACAAATAGAGAAGAAGTTT	1365
Db	1438	GATCTAAAGAAAAAATTAATGAAAAAATTTATACAGATATATAGGAAGAAAAATATTC	1497
Qy	1366	ATCAAGAAATCAAGAGAGAGATCAAAATTTAGAGAGAAAAATTTAGAGATGACACAAGAA	1425
Db	1498	ATTATATACATTTAAAAAACAATTTGATTTAGAAAGAAAAAACCTTAATGCACAAAAAGA	1557
Qy	1426	AGTTTCGAAGCCCGACGAAAAGCTTAACGATATCACTAAGAGATATGAAAAAGTCTG	1488
Db	1558	CAAAATTAATAAAATTTACTTGAAGATTATGAAGAACTCAAAAAAGGATTTATMAAATTACTT	1617
Qy	1486	AACGATCTATGATTTCCAAATTCACAATATACATGACCTGACCACTTCGAGAAATG	1548
Db	1618	GAAAAATTTTATGAATGAATTTTAATATAATTTTGCACAAAGATGCTGTAGATTAATA	1677
Qy	1546	ATGGAAAAACGCTACTCTTCCAAAGTGGAGAACTGCACACCATTAATACCTTTGCATCC	1605
Db	1678	TTCACTGCAAGATTAACATTAATATGTTGAAAAAACAAGATATATATTAATTTTCATCC	1737
Qy	1606	TATGAGATTTCAAGCATATCTTGAGAGCTCACAAAGCTTTAAGTATGAGAGAC	1665
Db	1738	TCTAATAATTTGTATATTAATGTTCAAAAAATTTAAAAAGGCTTTTCAATATCTTAAGAT	1797
Qy	1666	TATTTCTGCGCAACTTGTGTGGACAAAGACTAAAGTATTAACAGATCTCAATAGT	1725
Db	1798	TATTTCTTAACAAAAAGGATTTTCGAAAAGATTTTAATCATTTATTACTTTGAAACT	1857
Qy	1726	AAGATCGAAAAAGATCGAGACGCTTGTGAGACATTTAAGAAGATGAGAACACATTTG	1785
Db	1858	GGCCTCGAAGCTGATATTAATAAAATTTAACGAGAATAATTAAGATGAGTGAAGAAACAAAT	1917
Qy	1786	TTTGAGAGAGATTTACAAAAAGCAGAAATAAACACAGATGAGAAAGTCTCGAGCTCC	1845
Db	1918	CTAGAAAAAATTTTAAAGCTAACCATTCAGCAAAATG - - -CTTCTTAAAAATATCT	1974
Qy	1846	GATATTTGTAAGTCCCAAGTGCAGAGTGCCTCCTATGAACAAGATTAATCACTCAAG	1905
Db	1975	GATATTTGTAATAATTTACAAATGTAACAAAAAGTTTATTAATTAATAAAATTAAGAACCTTAAGA	2034
Qy	1906	AAGACTCACTCATTTCTGAAGAACGTGAGTTAAAAATATATATACATGTGCGGAATGT	1965
Db	2035	AAGATAGAAATTAATTTTAAAAAATGACACACTAAAGATAGTATTCATGTACCAATATTT	2094
Qy	1966	TATTAAGCAGGAGATTAAGCAGAGAACCTACTACTCATGTCTACTCAAGAAAGATTAAGC	2025
Db	2095	TATATAACCAAAATTAACACGAAACCATTAATTTATTTGTTATTAATAAAAGAAAGTAAAT	2154
Qy	2026	AAACGAAAGTTCATGCCCCAAAGTGCAGGCTCATCAAGAGAGAGAAAGAAATTT	2085
Db	2155	AAATTAAGAAATTTATTAACAAAAGTAAGAAACATGTTTAAAGAAACACMACTGTCTTA	2214
Qy	2086	AAAATGGAAGCAGTC - - - -AGATTAATCTCGAGCTTCCACAGAGAGAGATTAAC	2139
Db	2215	TCAAGTATTTACACAACCTTTTAGTTGACGACAGGAAACAACTGTAAGATGGGGGTCACTCC	2274
Qy	2140	GGACAGGCTTACCACCAAGCCCGGACACAAGGCCGTTCACTCTCGAAGCGCATAGCGTG	2199
Db	2275	ACACACACATTAATCCCAATCAGGAGAAACAGAAGTAAACGAAGAAACAGAAATTAACAGAA	2334
Qy	2200	CAAGCTCAAGCAGAGGACGAAGACAGGACAGCTTCACGTGCGACGTGCCGCTTCC - - -	2255

Db	2335	GAAACAGTAGGACACACACAAACGGTAAACATTAACCTTTACCCACCAACCAAGAAAGATCGACA	2394
Qy	2256	-----AAGGCTAAAGGCTCAAGTGCCTTACACCAACGACTCCTGTGAATTAAGACGCGAG	2310
Db	2395	CCAAAGAGAGTAAAGAGTTGTTGAAGAAATTCATATAGAAACATTAAGAGTATATGACATTTCCAA	2454
Qy	2311	AATGTCAAGCAAACTGGACCTACCTTGGAGAGCTCTATGAGTCCGTGAATACATCCTACATC	2370
Db	2455	GCCTTGACAAAAGACAGTTTATCTTAAGAAATAATGATGAATTTTAACTAAATACATATATA	2514
Qy	2371	TGCCCAAAATATATATCCCTCCCTCTCTCACAGCAGCTATGACAGAGAAATTTCTTAACAGTAC	2430
Db	2515	TGCTATTAATATATTTTATTTAGTATCAAACTGTAGATGACAAAAATTTTATAGAGATATAT	2574
Qy	2431	AAGATTAACCAAGAGAGAGAGAGTAACAGTGCCTCTGTGATCCACTGAGACTGCTGCTTC	2490
Db	2575	AATCTTACTCCAGAGAGAGAAAAAGATTTAAATCATGTGATCCATTGATTTTATTTATTT	2634
Qy	2491	AATATCAGAAACAACTTCCCGTTATGTATTTCTATGTTTCGATAGCCTCAGACATTTCTC	2550
Db	2635	AATATTCAAATATACATCACTGCTATGTATTCATTTATATGATATGATATGACAAATGATTTA	2694
Qy	2551	TTCTCACTGTTTCATGAGATATATGGAAGAGAGATGCTGCACACTGTATTAACCTCAA	2610
Db	2695	CAACATCTCTTTTGTGATTTATCTAAAGAAAGAAATGATTTATTTACATTAACCTAAA	2754
Qy	2611	GACACGACGAAGTTATAGAACCTTCGGAGG-----	2641
Db	2755	GAGAAATTCACATCAAAAATTTATTTAGAGAGACAAAACAAATACGTGAACATCATCT	2814
Qy	2642	-----AAGCTAAGAGAGGCTTCCACCTCTGTTAAACCTCTCTCTTC	2681
Db	2815	ACATCCAGTCCCTGGAAATTAACACCGTAATTACTGCTCAATCCGCAACACAGTAATTC	2874
Qy	2682	CAGTCCATGCAACCACTGTCTCTCACACCTCAAGCAA-----	2720
Db	2875	CAAAACCAAACTCAATGCGATCCTTACCAATACCCAAAATGGTGTAGCTGTATCATCT	2934
Qy	2721	-----GCCGGAAGGAGCGGTACGACGACACCTCTCACTGACCAACCTTAATATACGA	2775
Db	2935	GGTCTCTGTGTAGTTGAAGAAATCATGTATCCCTTTACAGTATTTGTCATTATGATGACAT	2994
Qy	2776	CTGAACCTGTTTGAAGACATCCTGTCTCCGCGAAGAAATAGAACAT---CTACCAAGAA	2832
Db	2995	TTGAAGAGTATGTGTAGTCTCTTAATCTCTGGAAATTAACATAAGTACCTATTCATTA	3054
Qy	2833	CTTATTTGGACAGAAATCTGCCGAACTTCTACGAGAGATATCTGAAGACAGCGACACA	2892
Db	3055	ACCATTTCTACAAACAGATGAGAAAAATTTTATGAGAAATTTTAAAAAATAATGATACC	3114
Qy	2893	TTCTATTAAGAGAGCTTCACTACTCTCGGAATCTAAAGCCGATGATATCACTCTT	2952
Db	3115	TATTTTATATATGATATCAACACATTCGTGTAATCTAATTTCAAAAGTAATTAACAGTTTG	3174
Qy	2953	AACGATTAATCTTAAGCTAAGAGCTGGAAGAGGACATCAATTAAGCTGAAGAACACTG	3012
Db	3175	ACC---GAACACAAAAAATGATTAATTAATGATGAATTAATAAATTTAAAAAGATACCTTA	3221
Qy	3013	CAACTGAGCTTGCAGCTGTACAACAGTACAACTGAACCTGGAAGAGCTCTTGACAG	3072
Db	3232	CAGTATCATTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3291
Qy	3073	AAGAGAGACAGTGGCAAGTATTAAGATGAGCATCAGAAAGTTGACTCTGCTCAAGAGCAG	3132
Db	3292	AAAAAGACACTGGCCAAAGCAAAATGCAAAATTTAAAAAACTTACTTTTATTAAGAACAA	3351
Qy	3133	CTTGAAGCAAACTCACTACTGATCAATCTCGAAACAGCTGCTGCAACATCTTCCAGTG	3192
Db	3352	TTAATAATCAAAATTTGATTTACTTATTAACCCACATATGATTTTCAAAAACCTTTTCTGT	3411
Qy	3193	TTCTTCAACAGAGAGAGGACCGGATGCGCGGAGACAGACACTCTGTGGAGACACC	3252
Db	3412	TTCTTTTAAACAAAAAGAGCTGAATATGCAAGAACTGAAACACCTTGTGAAGAACCA	3471

OY	3253	AAGATCTTCTCAAAACACTACAAGGCGCTGTCAAGTATTATAATGGCAGCTTCCTCT	3312
Dd	3472	AAAATATATTGAACACTTTATAAGGACTTGTAATATTATTAATGGTAACTCATCCCA	3531
OY	3313	CTGAAGACTCTCTCCGAGGAGACATCCAACCCGAGATTAAGTACCGCCGCTCGAGAC	3372
Dd	3532	TTAAAACTTTAAGTGAAGTATCATCTTAATCCAACGAAAGATTAATTBCCATTTAGAAAA	3591
OY	3373	TTCAAGGTCCTGTCTAAGCTCGAAGCCAGCTGAAGGACACCTGAACCTGGAGAAG	3432
Dd	3592	TTTAGAGCATTAAGTAAATAAGATGGAACAACTCAATGATTAATTTACATTTAGGAAGAA	3651
OY	3433	AAGCTCAAGCTACCTCTACGAGCTGCACACCTCATGCTGCCGAGCTCAAGAACTCTT	3492
Dd	3652	AAATTAATCTTCTATTAAGTGAAGTATTCATCATTTTAATAGTAATTTAAACAAAGTAATA	3711
OY	3493	AAGAACAAGAACTACACCGGCAATAGCCCCAAGCGAATAAATACAGAGCTGAATTAAGCA	3552
Dd	3712	AAAAATAAATAATTATACAGTAATTTCTCCAGTGAAGAAATAATAGAAAGTTAAGCAACT	3771
OY	3553	CTGGAATCTTACAAGAAGTTCTGCTTGAAGAACAGATGTGC-----	3596
Dd	3772	TTAAAAATCTTACGAAATTTTTTCCAGAAAGCAAAGTTACAAAGTTGTAATCTCACCT	3831
OY	3597	-----CACGTGGTGTCTGAATGTGGCTCCGACACAGCTGGACACTGTCA	3641
Dd	3832	CACACAGATGTAACTCCATCTCCATTAATCTGTAAAGGTAAAGTGTAAGTTCACAGATCCACA	3891
OY	3642	ACCTAAGAA-----GCCGTGATCTACTCATGTGCGAGCCGAGTCCATTAATTAAC	3693
Dd	3892	AAGAAGAAACACAAATTAACCACTTCAGGCTCTTTATTAACGAATTAACAACTAGTA	3951
OY	3694	ACATCTCAGAACGTCGACGATTAAGTGAAGTCGATAGTCATATTTGTGCTATCTTGGCAG	3753
Dd	3952	CAATTACAAATTAATGACGAAGAATGATCTCTTAAGTTGTATTAACCATTTTTGGAGAA	4011
OY	3754	AGCGAGGAGACTCGATGTGACCTCGGCCAGAGTGTGTCACCGGTGAGAGCTCACTCTCC	3813
Dd	4012	TCCGAGATTAATGACGAATTTTAAGATCAAGTAGTAGTAACGGGAAGC---AATATCTGTC	4068
OY	3814	GTCATTTGATTAACATTTCTGTCCAAAATCGAGAACGAATACGAAGTGTCTATCTGAACCT	3873
Dd	4069	ACAAATGATAATTAATCTCTCAGGATTTGAAAATGAATTAAGATTAATTAATTAACCT	4128
OY	3874	CTGGCAGGCGCTATAGTGTCTCAAGAAAACAGCTGGAATAATACGTGACCTTCAT	3933
Dd	4129	TTACTGTGAGTATTAAGAACCTTAATAAAAAACAATTGAAAAAACATTTATACATTTAAT	4188
OY	3934	GTCAACTGAAGACATTTCTGAACAGCCGCTTAATTAAGAGAAAAATTTCAAGACGTC	3993
Dd	4189	TTAAATTTGAACGATATCTTAATTAATCAAGCTTAAGAAACGAAATATTTCTTAAGNTGA	4248
OY	3994	TTGGAGGCGACTGATTTCCCTTAATAAGACCTGACCTCTCAACACAGTTGTCAAGAC	4053
Dd	4249	TTAGAAATCTGATTTAATGCAATTTAAACATATATCTCAATGAATACATTTTAAGAT	4308
OY	4054	CCATPACAAGTCTCTCAATAAAGAGAGAGGATTAATTTCTGTAGTTACAATATATC	4113
Dd	4309	TCATTTAAATTAATTAATGATTCAGAACAAAAAACACCTTTTAAAAAGTTACAAATATATA	4368
OY	4114	AAGAACTCCACGACACCGATATCAATTTCCGAATGATGTGCTGGGATTAACAAATC	4173
Dd	4369	AAAGAAATCAGTAGAAATTAATTAATTTGACACAGGAAGTATTAAGTTAATGAAG	4428
OY	4174	CTGAGCAAAAATTCAGTCTGACCTTGACTGTCAATTAAGATAATATACAGATTAAG---	4230
Dd	4429	GTTTATAGGAATTAATTAAGATTAATTAAGATTAATTAATTAATTAATTAATTAATTAATTA	4488
OY	4231	-----CAA	4233
Dd	4489	GAGAGATTTCCATATACCAACCAACACTCGTCCAGCAAAAAACAGACGAACAA	4548

OY	4234	GGCGAATGAAATATCTGCGCTTCGATTAACATCGAAACCTGCAAGACAGG	4293
Db	4549	AAGAAAGAAAGTAAATTTCTTCATTTTAAACAAACATTTGAGACCTTTATACATTA	4608
OY	4294	AACGACAAATGACCTCTTCTGTAATTCACCTGAGGCGCAAGTCTCACTATCTTAC	4353
Db	4609	GTTAATTAATTTGACGATTACTTAATTAACCTAAAGGCAAGATTAACGATTTGTAATG	4668
OY	4354	GAGAAAGACATGTGGAGTTAAATCAAGAGCTGGAACCTACCTAAAAACAATCCAAGC	4413
Db	4659	GAAAAAGATGAAGCACATGTTAAATTAACCTAACTAGTATTTAAAAAGCAATTTATGAC	4728
OY	4414	AAGCTGGCAGATTTTAAAGAAAAATTAACAATTTGCTGCGAATGACAGACCTGCTAC	4473
Db	4729	AAAATATGATCTTTTAAACAAACCTAACGACCTTGAAGCAATTTAAAAATTTGTAATGAT	4788
OY	4474	TATTAACACACAAATCTCTCTGACAGATTTCTGTCACCTGGCATGGATGGTTGCAAAACCTC	4533
Db	4789	GATACGAAAAAAGATATGCTTGGCAAAATTAATAGTACAGAGTTAG--TTGCAAAATTTT	4845
OY	4534	GCCAAACACAGTCTGAGCAATTCGTCGACGGCAACCTGACGGGATCTGACATCTCC	4593
Db	4846	CCTAATACATTAATTCACAAATTAATGGAAGAAAAATTCACAGATATGTTAAACATTTCA	4905
OY	4594	CAGCACAATGGGTGAGAAAGAACAGTCCGCCATGATGAGGCTGTTTACAGGATCTGGAC	4653
Db	4906	CACACCAATGCGCTAAAAAACAATGTCAGAAAAATTTCTGATGTTTCCAGACATTTAGAT	4965
OY	4654	GAGCGCGAAGATGCAAGTGTCTCTGACATCAACAACAAGAAAGAAATTAAGTGTGGAG	4713
Db	4966	GAAAGAGAAAGATGTAATATGTTTATTAATTAATACAAACAAGAGTGATTAATATGTTGAA	5025
OY	4714	AACCCAAACCTTACTGCAATGAAAAACAATGCGGGGTGTGACGCGGATGCTAATGACAC	4773
Db	5026	AATCCAAATCCCTACCTGTAACGAAAAATTAATGATGATGATGATGATGATGATGATGAT	5085
OY	4774	GAGGAAGACAGGGGTCTTAACGAGAAAGAAATACATGAGTGAATGAGTAAAGCCGACATCC	4833
Db	5086	GAAAGAGATTTGAGTACGATGACGACGAGAAAGAAATACATGATGATGATGATGATGAT	5145
OY	4834	TATCCACTCTTCGACGAGGATTTTGTCTCCACCTCAATATTCCTGGGATCTCTCTCTG	4893
Db	5146	TATCCACTTTTGATGATGATATTTTTCGAGTTCCTCTCACTCTTGAAGAAATATCATTTCTTA	5205
OY	4894	CTGATCTCTTCGATCTCTGATCTGTACAGCTTCATCTTAATGATGATG	4939
Db	5206	TTAATATCATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5251
RESULT 13			
PFAM5AA			
LOCUS		5754 bp	linear
DEFINITION		P.falciparum major merozoite surface antigen (PFMSA) mRNA, complete cds, isolate FC27.	
ACCESSION		M19143	
VERSION		M19143.1	GI:160412
KEYWORDS		antigen; glycoprotein.	
SOURCE		P.falciparum [isolate FC27 from Papua New Guinea], cDNA to mRNA, clone Ag75, g1.i, g126, pEBG3.3.	
ORGANISM		Plasmodium falciparum	
REFERENCE		Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		1 (bases 1 to 5754) Peterson, M.G., Coppel, R.L., McIntyre, P., Langford, C.J., Woodrow, G., Brown, G.V., Anders, R.F. and Kemp, D.J.	
TITLE		Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum	
JOURNAL		Mol. Biochem. Parasitol. 27 (2-3), 291-301 (1988)	
MEDLINE		88142999	
PUBMED		2449612	
FEATURES		Location/Qualifiers	
source		1..5754 /organism="Plasmodium falciparum" /isolate="FC27 (Papua, New Guinea)"	

```

gene
  /db_xref="taxon:5833"
  1..5754
  /gene="PmmsA"
  <1..5754
  /gene="PmmsA"
  121..526
  /gene="PmmsA"
  /note="precursor"
  /codon_start=1
  /product="major merozoite surface antigen"
  /protein_id="AA029653.1"
  /db_xref="GI:160413"
  /translation="MKIIIFPLCSFLFIINOCVTHESGYELKLEALDELAVLTGYS
  LFOEKAVLNGSTGTAVTATSTPGSSGVSAGSAVASAASGSGSVAGSGS
  RRTNPDSNSDSNRTKYADLKHRYQNTLFTIKELYPELPULTLNMHLISNVGGR
  LLDYEEINELLKLYFLFYDDLRAKLNDACNSYCOQIPFNKIRANEDLVKLTFFG
  RLPDNDIKDNYGMKMEYIKKRTTIANINLEISCKTIDONKNADNEGKKLYQAA
  YNLFLYNGLOEAMHLISVLEKRIDITLKNBNIKKLEIDIKITDAENPTGSPN
  LPEKKKEVEGHEERKEIKELATIKENIDSLTDPLELEYLRKKNKADVPKSODP
  KSVQIPKVPYNGIVYPLPLDINHSLAANDKSYGDLMDPDKRINEKILINDPK
  RKITINNIKOIDLEBKINNTKQNKCLLEDYEKSKDIYELLERPEAKFPNNPDR
  DVAVRIFSAITYAVEKORINNKFSNSNVSYNVQKALKALSYLEDLSLRGISKD
  NHYETLRGLFADIRKLTLEIKSSENKLTLENFGSLTHSAASLEVSIVLQVQV
  LIKLETLRKLEIFLKAQLOKDSIHVPNIPQPKPEBYUUYLVKEDVKEIPK
  KDMKLEQVASYSTOPLVASSETDEGSHTHLDSGSETEVEETFEVGHGTTH
  TVTILPLKEESAREKVVANSTIEHNSDLSOLTKVYKLTKEIDELFTKSYICHY
  LVSSNSMDOKLLEYNLTPEEENELKSDPDLPLFNQNNIPAMYSLYDSNNIDQHL
  FFEYLQEMTYLHRLKEENHKKLLEBOKOITGTSSTSPGNTVTVAQSATSNS
  NQOSNASSTNTQNGVAVSSGPAVVEESHDPPLTVSISNDLGIYSLNIGKRTVPS
  LTISTTEMEKYENILKNNDYFPNDIDQFVKSNSKYTGLEOKNALNDEIKKJ
  TLOSPFLYNNKRLKLDLFPKKKELGODKKQIKLTKYKOLESKLNNPNHVLCD
  NSEVFPKKKEAEIAETENTLENRILKHKGLVYKNGSSPLKLTISEVSIQDE
  YANLEFRALSKIDGKLDNLDLKKKLSIFSSGHLITELKEVYIKKNTQNSPDS
  NNRVNEALKSYENLPEPAKVTYVTPQDPVTSVPSVRSGSGSTKEIQTPS
  SLELDQOVOLQWDEEDSLVVLPTGEENDEYLDVVTGEALSYVDNLLSGE
  ENEDVYILKPLAGVYRSLSKQIKENITFPLNDINLSLRKREYLDVLESDLMQ
  FKHSISNEYIIEDESKLNSOQNTLKSXYIKYIESVDNIDKFAQESYIEYLAK
  KDLIESIKYVIEKEERKPPSPPTTPSPADTDQKRSKRLPLTINLTLYNNLVNN
  IDYLLINLAKIINDCNVEKDAHYKIRLSLKAIDKIDIDFKNTNDEPAIRKLIND
  TKRMGLKLTSTGLVQNPNTIISKLLEGRQDMINISQHKVQCKQESNGCRHLL
  ERECKKCLANQKGDGCVENPNPENNNGGADACVTEBDGSSSRKLTICBCTCRK
  DSYPLPDGIFCSCSSNFIISFLILMLILYFSI"
  121..177
  /gene="PmmsA"
  178..5223
  /gene="PmmsA"
  /product="major merozoite surface antigen"
  /protein_id="PmmsA"
  /product="major merozoite surface antigen"
  /db_xref="GI:160413"
  /translation="MKIIIFPLCSFLFIINOCVTHESGYELKLEALDELAVLTGYS
  LFOEKAVLNGSTGTAVTATSTPGSSGVSAGSAVASAASGSGSVAGSGS
  RRTNPDSNSDSNRTKYADLKHRYQNTLFTIKELYPELPULTLNMHLISNVGGR
  LLDYEEINELLKLYFLFYDDLRAKLNDACNSYCOQIPFNKIRANEDLVKLTFFG
  RLPDNDIKDNYGMKMEYIKKRTTIANINLEISCKTIDONKNADNEGKKLYQAA
  YNLFLYNGLOEAMHLISVLEKRIDITLKNBNIKKLEIDIKITDAENPTGSPN
  LPEKKKEVEGHEERKEIKELATIKENIDSLTDPLELEYLRKKNKADVPKSODP
  KSVQIPKVPYNGIVYPLPLDINHSLAANDKSYGDLMDPDKRINEKILINDPK
  RKITINNIKOIDLEBKINNTKQNKCLLEDYEKSKDIYELLERPEAKFPNNPDR
  DVAVRIFSAITYAVEKORINNKFSNSNVSYNVQKALKALSYLEDLSLRGISKD
  NHYETLRGLFADIRKLTLEIKSSENKLTLENFGSLTHSAASLEVSIVLQVQV
  LIKLETLRKLEIFLKAQLOKDSIHVPNIPQPKPEBYUUYLVKEDVKEIPK
  KDMKLEQVASYSTOPLVASSETDEGSHTHLDSGSETEVEETFEVGHGTTH
  TVTILPLKEESAREKVVANSTIEHNSDLSOLTKVYKLTKEIDELFTKSYICHY
  LVSSNSMDOKLLEYNLTPEEENELKSDPDLPLFNQNNIPAMYSLYDSNNIDQHL
  FFEYLQEMTYLHRLKEENHKKLLEBOKOITGTSSTSPGNTVTVAQSATSNS
  NQOSNASSTNTQNGVAVSSGPAVVEESHDPPLTVSISNDLGIYSLNIGKRTVPS
  LTISTTEMEKYENILKNNDYFPNDIDQFVKSNSKYTGLEOKNALNDEIKKJ
  TLOSPFLYNNKRLKLDLFPKKKELGODKKQIKLTKYKOLESKLNNPNHVLCD
  NSEVFPKKKEAEIAETENTLENRILKHKGLVYKNGSSPLKLTISEVSIQDE
  YANLEFRALSKIDGKLDNLDLKKKLSIFSSGHLITELKEVYIKKNTQNSPDS
  NNRVNEALKSYENLPEPAKVTYVTPQDPVTSVPSVRSGSGSTKEIQTPS
  SLELDQOVOLQWDEEDSLVVLPTGEENDEYLDVVTGEALSYVDNLLSGE
  ENEDVYILKPLAGVYRSLSKQIKENITFPLNDINLSLRKREYLDVLESDLMQ
  FKHSISNEYIIEDESKLNSOQNTLKSXYIKYIESVDNIDKFAQESYIEYLAK
  KDLIESIKYVIEKEERKPPSPPTTPSPADTDQKRSKRLPLTINLTLYNNLVNN
  IDYLLINLAKIINDCNVEKDAHYKIRLSLKAIDKIDIDFKNTNDEPAIRKLIND
  TKRMGLKLTSTGLVQNPNTIISKLLEGRQDMINISQHKVQCKQESNGCRHLL
  ERECKKCLANQKGDGCVENPNPENNNGGADACVTEBDGSSSRKLTICBCTCRK
  DSYPLPDGIFCSCSSNFIISFLILMLILYFSI"
  121..177
  /gene="PmmsA"
  178..5223
  /gene="PmmsA"
  /product="major merozoite surface antigen"
  /protein_id="PmmsA"
  /product="major merozoite surface antigen"
  /db_xref="GI:160413"
  /translation="MKIIIFPLCSFLFIINOCVTHESGYELKLEALDELAVLTGYS
  LFOEKAVLNGSTGTAVTATSTPGSSGVSAGSAVASAASGSGSVAGSGS
  RRTNPDSNSDSNRTKYADLKHRYQNTLFTIKELYPELPULTLNMHLISNVGGR
  LLDYEEINELLKLYFLFYDDLRAKLNDACNSYCOQIPFNKIRANEDLVKLTFFG
  RLPDNDIKDNYGMKMEYIKKRTTIANINLEISCKTIDONKNADNEGKKLYQAA
  YNLFLYNGLOEAMHLISVLEKRIDITLKNBNIKKLEIDIKITDAENPTGSPN
  LPEKKKEVEGHEERKEIKELATIKENIDSLTDPLELEYLRKKNKADVPKSODP
  KSVQIPKVPYNGIVYPLPLDINHSLAANDKSYGDLMDPDKRINEKILINDPK
  RKITINNIKOIDLEBKINNTKQNKCLLEDYEKSKDIYELLERPEAKFPNNPDR
  DVAVRIFSAITYAVEKORINNKFSNSNVSYNVQKALKALSYLEDLSLRGISKD
  NHYETLRGLFADIRKLTLEIKSSENKLTLENFGSLTHSAASLEVSIVLQVQV
  LIKLETLRKLEIFLKAQLOKDSIHVPNIPQPKPEBYUUYLVKEDVKEIPK
  KDMKLEQVASYSTOPLVASSETDEGSHTHLDSGSETEVEETFEVGHGTTH
  TVTILPLKEESAREKVVANSTIEHNSDLSOLTKVYKLTKEIDELFTKSYICHY
  LVSSNSMDOKLLEYNLTPEEENELKSDPDLPLFNQNNIPAMYSLYDSNNIDQHL
  FFEYLQEMTYLHRLKEENHKKLLEBOKOITGTSSTSPGNTVTVAQSATSNS
  NQOSNASSTNTQNGVAVSSGPAVVEESHDPPLTVSISNDLGIYSLNIGKRTVPS
  LTISTTEMEKYENILKNNDYFPNDIDQFVKSNSKYTGLEOKNALNDEIKKJ
  TLOSPFLYNNKRLKLDLFPKKKELGODKKQIKLTKYKOLESKLNNPNHVLCD
  NSEVFPKKKEAEIAETENTLENRILKHKGLVYKNGSSPLKLTISEVSIQDE
  YANLEFRALSKIDGKLDNLDLKKKLSIFSSGHLITELKEVYIKKNTQNSPDS
  NNRVNEALKSYENLPEPAKVTYVTPQDPVTSVPSVRSGSGSTKEIQTPS
  SLELDQOVOLQWDEEDSLVVLPTGEENDEYLDVVTGEALSYVDNLLSGE
  ENEDVYILKPLAGVYRSLSKQIKENITFPLNDINLSLRKREYLDVLESDLMQ
  FKHSISNEYIIEDESKLNSOQNTLKSXYIKYIESVDNIDKFAQESYIEYLAK
  KDLIESIKYVIEKEERKPPSPPTTPSPADTDQKRSKRLPLTINLTLYNNLVNN
  IDYLLINLAKIINDCNVEKDAHYKIRLSLKAIDKIDIDFKNTNDEPAIRKLIND
  TKRMGLKLTSTGLVQNPNTIISKLLEGRQDMINISQHKVQCKQESNGCRHLL
  ERECKKCLANQKGDGCVENPNPENNNGGADACVTEBDGSSSRKLTICBCTCRK
  DSYPLPDGIFCSCSSNFIISFLILMLILYFSI"
  121..177
  /gene="PmmsA"
  178..5223
  /gene="PmmsA"
  /product="major merozoite surface antigen"
  /protein_id="PmmsA"
  /product="major merozoite surface antigen"
  /db_xref="GI:160413"
  /translation="MKIIIFPLCSFLFIINOCVTHESGYELKLEALDELAVLTGYS
  LFOEKAVLNGSTGTAVTATSTPGSSGVSAGSAVASAASGSGSVAGSGS
  RRTNPDSNSDSNRTKYADLKHRYQNTLFTIKELYPELPULTLNMHLISNVGGR
  LLDYEEINELLKLYFLFYDDLRAKLNDACNSYCOQIPFNKIRANEDLVKLTFFG
  RLPDNDIKDNYGMKMEYIKKRTTIANINLEISCKTIDONKNADNEGKKLYQAA
  YNLFLYNGLOEAMHLISVLEKRIDITLKNBNIKKLEIDIKITDAENPTGSPN
  LPEKKKEVEGHEERKEIKELATIKENIDSLTDPLELEYLRKKNKADVPKSODP
  KSVQIPKVPYNGIVYPLPLDINHSLAANDKSYGDLMDPDKRINEKILINDPK
  RKITINNIKOIDLEBKINNTKQNKCLLEDYEKSKDIYELLERPEAKFPNNPDR
  DVAVRIFSAITYAVEKORINNKFSNSNVSYNVQKALKALSYLEDLSLRGISKD
  NHYETLRGLFADIRKLTLEIKSSENKLTLENFGSLTHSAASLEVSIVLQVQV
  LIKLETLRKLEIFLKAQLOKDSIHVPNIPQPKPEBYUUYLVKEDVKEIPK
  KDMKLEQVASYSTOPLVASSETDEGSHTHLDSGSETEVEETFEVGHGTTH
  TVTILPLKEESAREKVVANSTIEHNSDLSOLTKVYKLTKEIDELFTKSYICHY
  LVSSNSMDOKLLEYNLTPEEENELKSDPDLPLFNQNNIPAMYSLYDSNNIDQHL
  FFEYLQEMTYLHRLKEENHKKLLEBOKOITGTSSTSPGNTVTVAQSATSNS
  NQOSNASSTNTQNGVAVSSGPAVVEESHDPPLTVSISNDLGIYSLNIGKRTVPS
  LTISTTEMEKYENILKNNDYFPNDIDQFVKSNSKYTGLEOKNALNDEIKKJ
  TLOSPFLYNNKRLKLDLFPKKKELGODKKQIKLTKYKOLESKLNNPNHVLCD
  NSEVFPKKKEAEIAETENTLENRILKHKGLVYKNGSSPLKLTISEVSIQDE
  YANLEFRALSKIDGKLDNLDLKKKLSIFSSGHLITELKEVYIKKNTQNSPDS
  NNRVNEALKSYENLPEPAKVTYVTPQDPVTSVPSVRSGSGSTKEIQTPS
  SLELDQOVOLQWDEEDSLVVLPTGEENDEYLDVVTGEALSYVDNLLSGE
  ENEDVYILKPLAGVYRSLSKQIKENITFPLNDINLSLRKREYLDVLESDLMQ
  FKHSISNEYIIEDESKLNSOQNTLKSXYIKYIESVDNIDKFAQESYIEYLAK
  KDLIESIKYVIEKEERKPPSPPTTPSPADTDQKRSKRLPLTINLTLYNNLVNN
  IDYLLINLAKIINDCNVEKDAHY
```


Db 409 GGTCCAAAGTGTCCAAAGTGTACAAAGTCCATCTCTTCAAAAGCACTTACCTGTTCA 468
QY 325 GGGAAACAGTCGAAAGAAACCAATCCATCTGACACTGATGGATTCGAGCCGCAAGTCTCC 384
Db 469 AATACCTTCATCTGGTGCAGGCTCCAGCTGATGCAAGGATTCAGATCTAAATCTTAC 528
QY 385 GCCGACCTCCAGCAGCCAGTGAAGAACTATCTCTCATCTATCATAGAGGATGGAATGCCA 444
Db 529 GGTGATTTTAAACACAGAGATAGCAAAATTTACTTGTTCATATTAAGAAGTCAAAATGCC 588
QY 445 CAGTGTTCGACCTCATATCATATGCTGACACTGTGTGATTAACATTTATGCTTCAAA 504
Db 589 GAACCTTTGATTTAAACCATATGTTAACTTGTGTGATTAATTAATGATGTTTCAAA 648
QY 505 TATTCATTTAGAGGTTAGCAAGAGATGCAATGCACTCTGACAAAGTGAATTTCTCTTC 564
Db 649 TATTTAATTTAGATGATGATGAAATTAATGAAATTAATTAATTAATTAATTTATTTT 708
QY 565 GACTTGTAAAGGCCAAACTGATGACGTTTGCCCAATGACTATTTGCAAAATTCATTC 624
Db 709 GATTTATTAAGCAAAATTTAAATGATGATGATGATGATGATGATGATGATGATGATG 768
QY 625 AATTTGAGATCAGAGCCAGCAAGGTTGACGTTTGAAGAAGTGTCTTCGATATGCC 684
Db 769 AATCTTAAATTCGTGCAATGATTAATGACGTTTAACTTAACTTGTGATTTAGA 828
QY 685 AAGCCTCTGACACATCTCAAGACATGTGGGAAAGATGGAAGATTAATTAATAAAGAT 744
Db 829 AATCATTAGACAAATTTAAAGATATGTAGGAAATGGAAGATTAATTAATAAAGAT 888
QY 745 AAGAGACCATGAGAACATTAACGAGTGTATGGAAGATCCAAAAGCATTAGCAAA 804
Db 889 AAAACACCATAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 948
QY 805 AATTAAGATTCACACAGAGAGAAAGAAAGAAAGTGTGACAGGCCGATGACATG 864
Db 949 AATTAAGATTCACAGATTAATTAAGAGAGAAAGAAAGTGTGACAGGCCGATGACATG 1008
QY 865 TCCATCTATTAACAAACAGCTTGAAGAGCCATTAACCTCATGACGCTACTGGAAGCCG 924
Db 1009 TCTATTTCAATTAACAAATTAAGAAAGCAATTAATTAATTAATTAATTAATTAATTA 1068
QY 925 ATAGACACCTTCAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 984
Db 1069 ATTGAACCTTAAAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGTGAAGAAAGT 1128
QY 985 AAGAACTCTCCGCGCAACTCTGGGAAACCCCTAACAAGCTGTGACAGAGCAAG 1044
Db 1129 AAAAATCCCCACCGGCAATTTCTGGAAATACACCAATTAATCTCTTGTATTAAGAA 1188
QY 1045 AAGATTAAGAGACAGAAAGATCAAGAGATGCGCAAAACATTAATTAATTAATTAAT 1104
Db 1189 AAAATTCGAGGAAACGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1248
QY 1105 GATTCCTCTTACATGATCCCTTGTGAGTGAAGTACTACTGAGAGAGAAAGTAAAGAT 1164
Db 1249 GATGATTTATTTACTGATCCACTGATTAATTAATTAATTAATTAATTAATTAATTA 1308
QY 1165 ATAGACATCTCCGCAAGTGA-----GACAAAGAAATCAACCGAATCTTAATGA 1215
Db 1309 GTTGAATTAACACCTTAATTAACAGAAATCTGTTCAATTAATTAATTAATTAATTA 1368
QY 1216 TATCCCAATGTGTGACATGCTGTCTTATTAACGATTAACAAAGCTCT----- 1268
Db 1369 TATCCCAATGTGTGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1428
QY 1269 --CAAGAGCTCAATAGCTTGTGATGATTAATTAATTAATTAATTAATTAATTAAT 1326
Db 1429 GATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1488
QY 1327 TCTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1386
Db 1489 GAAAAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1548

QY 1387 ATCAAAATTTAGAGAGAGAAATTTAGAGTGAAGAAAGTGTAGAGACCGACAGCAAA 1446
Db 1549 ATTGATTTTGAAGAAAGAAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1608
QY 1447 AGCTTAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1506
Db 1609 GATTTAAGAAAGTCAAAAGAAAGTAAAGAAATTAATTAATTAATTAATTAATTA 1668
QY 1507 TTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1566
Db 1669 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1728
QY 1567 AAAAGTGAAGAACTGACACACCATTAATTAATTAATTAATTAATTAATTAATTAAT 1626
Db 1729 AATGTTGAAGAAAGAAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1788
QY 1627 CTGAGAGCTTACCAAGCTCTTAAGTATTAAGAGACTATTTCTGCGGAACATGTT 1686
Db 1789 GTTCAAAATTTAAAGAAAGCTCTTCTCATATCTGGAAGATTAATTAATTAAGAAAG 1848
QY 1687 GTGAGAGAAAGTAAAGTAAAGATTAATTAATTAATTAATTAATTAATTAATTA 1746
Db 1849 TCTGAAAAAGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1908
QY 1747 ACGCTTGTGAGAACTTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1806
Db 1909 AATTTAAGCAAGAAATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1968
QY 1807 GACGAAGTAAAGCAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 1866
Db 1969 CTACACATTAAGCAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 2025
QY 1867 CAGAAAGTCTCTCATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1926
Db 2026 CAAAAAGTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2085
QY 1927 AACGTGAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1986
Db 2086 AATGCAAACTTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2145
QY 1987 GAACATCTACCTCATCTGACTCAAGAAAGATTAAGAGATTAAGAGATTAAGAGAT 2046
Db 2146 GAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2205
QY 2047 AAGTGAAGCTGTATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 2102
Db 2206 AAGTGAAGCTGTATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 2265
QY 2103 --AGATTAAGCTGTATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 2160
Db 2266 GTTGAAGAGAGAAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 2225
QY 2161 GGAACAGAGCGCTTCACTGATGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 2220
Db 2326 GGAAGAAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTA 2385
QY 2221 AAGCAGC-----ACAGCTTCAAGTCCAGTGTGCTTCAAGAGCTTAAGCTCA 2271
Db 2386 ACGGTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2445
QY 2272 GTGCCAAGCAGCAGCTCTGTGATTAATTAATTAATTAATTAATTAATTAATTAAT 2331
Db 2446 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2505
QY 2332 CTGAGAGAGCTTATGATTTCTGATTAATTAATTAATTAATTAATTAATTAATTAAT 2391
Db 2506 CTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2565
QY 2392 TCTCAGAGCACTTAAGAGAGATTTTAAGAGATTAAGAGATTAAGAGATTAAGAG 2451
Db 2566 TCAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2625

QY 2452 AGTAACCTGCTCTTGTGATCACTGAGCCTGCTGTTCAATATCCAGAACACATTCCTC 2511
DB 2626 AATGAATTAATAATCATGATCCATATGATTTATTTATTTAATATTCACAAAATAACATTAACCT 2685
QY 2512 GTTATGATTTTGTATGTTGATGACCTCAGACATTCCTCTCTCAACTGTTCAATGAGATTA 2571
DB 2686 GCTATGATTTATATATGATGATGATGACATGATTTTCAACATCTCTTTTGTAAATTA 2745
QY 2572 TATGAGAGAGATGCTGACCTGTATTAACATCAAGACAAACGACAGATTAAGAC 2631
DB 2746 TATCAAAAAGAAATGATTTATTTATTTACATTAACATAAGAGAAATACATCAAAAAA 2805
QY 2632 CTTCGAGAGAGCTTAAGAGGCTCTC 2657
DB 2806 TTATTTAGAGGACAAACAAATACTGGAACATCATCTACATCCAGCTCGAATAACA 2865
QY 2658 -----CACCTGTTAAACCTCTCTTCAGAGCTCAATGCAACCTGTC 2702
DB 2866 ACCGTAATTAATGCTCAATTCGCAACTCAGATTAATCCCAAAACCAATCAATGCA 2925
QY 2703 TCTCAGACCTCAAGACA-----GCCGAAGTGAGGCT 2736
DB 2926 TCTCTTACCAATACCCAAATGATGCTGATCATCTGCTGCTGATGTTGAAGAA 2985
QY 2737 AACGAGACACCTCTCACTCGACCAACCTTAATACCTCACTGAAACCTTTGAGAACATC 2796
DB 2986 AGCATGATGCTCTTAACGATATGCTATTTAGTAACGATTTGAAGATTTGTTGCTC 3045
QY 2797 CTGCTCTCGGCAAGATTAAGACAT---CTACCAAGACTTATTTGACAGAAATCGTCC 2853
DB 3046 TTTAATCTGGAATTAACAACTAAAGTACCTATTCATTAACCATTTCTACAAACAGATG 3105
QY 2854 GAGAACTCTACGAGAGATCTGAAAGACAGCAGACATCTATTAAGAGAGCTCACT 2913
DB 3106 GAAAAATTTTATGAGAAATATTTTAAATAATATGATTAATCTATTTTAAATGATATCTAAA 3165
QY 2914 AACTTCGTGAATCTAAAGCCGATGATATCACTCTCTTAACGATGAATCTAAACGTAAAG 2973
DB 3166 CAATTCGTAATCTAATTTCAAAAGTATTAACAGGTTTGACC---GAAACACAAAAAAT 3222
QY 2974 AAGCTGGAAGAGACATCAATTAAGCTGAAGAGACACTGCACTGAGCTTGACCTGTAC 3033
DB 3223 GCATTTAATGATGAATTTAAAAATTTAAAGATCTTTCAGTTATCTATTTGATTTAAT 3282
QY 3034 AACAACTCAACACTGAACCTGGAGAGCTCTTCGACAGAGAAAGACAGCTGCAAGAT 3093
DB 3283 AATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3342
QY 3094 AAGATGAGATCAAGAGTTGACTGCTGCTCAAGGAGAGCTTGAAGCAAACTCAACTCA 3153
DB 3343 AAAATGCAAAATTAATAAACTTATTAATAAGAACAACTTAAGAAATTTGATTTCA 3402
QY 3154 CTGAACAATCCGAACAGCTACTGACAGACTTCTGAGTCTTCAACAGAGAGAA 3213
DB 3403 CTTAATTAACCCCAATTAATGATTAACAAAACCTTTCTGTTTCTTAACAAAAAAGAA 3462
QY 3214 GCCGAGATCCGCCGAGAGAGACACTGAGAACACACACATCTCTCAACACATAC 3273
DB 3463 GCTAATATAGCAAACTGAAACACACATTAAGAAACACAAAATTAATTAAGAAACATTAAT 3522
QY 3274 AAAGGCTCGTCAAGTATTAATGAGGAGTCTTCTCTCTGAGAGACTCTCCGAGAG 3333
DB 3523 AAAGGACTTGTTAATATTAATGATGATCATCTCATTAATAAACTTAAGGAACTA 3582
QY 3334 AGCATCCAGACCGAGATTAATAGCCAGCTCGAAGAACTTCAAGGCTCTCTTAAGCTC 3393
DB 3383 TCAATTTCAAAAGAGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3642
QY 3394 GAAGGCAAGCTGAAGACAACTGAGACTGAGAGAGAAAGAGAGCTAGCTACTCTCTAGC 3453
DB 3643 GATGAGAAACCTCAATGATTAATTTACATTTAGGAAAGAAATTAATTTCTTTTATCAAGT 3702
QY 3454 GGAATGATCACTGATGCTGAGCTCAAGAGATCAATTAAGAAACAGACATCAACCGC 3513

DB 3703 GGATTAATCAATTTATTAATGAAATTAATAAGAGTAATAAAAAATTAATTAACAGCT 3762
QY 3514 AATGACCCAGAGAGATTAATTAACAGCTGATTAATTAACAGCTGATTTCAAGAGCTTC 3573
DB 3763 AATTCCTCAAGAGAAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTT 3822
QY 3574 CTGCTGAAGGAGAGATGCTGC-----CACTGT 3602
DB 3823 CTCCAGAGCAAAAGTTTACACAGCTTTAACTCCACTCAACAGATTAATCTCATCT 3882
QY 3603 GGTGCTGAATCTGCTCCGACACACTGAGCAGCTCTCACTTAAGAA-----GCCT 3654
DB 3883 CCATTAATCTGAAGGTAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3942
QY 3655 GCATTAATCAATGCTGAGGAGCCGAGCTCAATTAATTAACATCTCAACAGCTGAGCAT 3714
DB 3943 ACTTCAGGCTCTTAATTAATTAACAGAAATTAACAGAGATTAACAAATTAATTAACGAA 4002
QY 3715 GAGGTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3774
DB 4003 GAGAT 4062
QY 3775 CTGCGCAGGTGCTGACCCGATGAGGCTGCTCACTCTCTGATGATGATTAATTAATCTGCTC 3834
DB 4063 TTAGATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4119
QY 3835 AAATTCGAGAGATTAACAGAGCTCTATCTGAACCTCTGCAAGGCTCTATAGCTCT 3894
DB 4120 GGATTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4179
QY 3895 CTCAAGAAACAGCTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3954
DB 4180 TTTAAAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4239
QY 3955 AACAGCCGCTTTAATTAAGAGAAATTTCAAGAACTGTTGAGAGGAGCTGATTC 4014
DB 4240 AATTCAGCTTTAAGAAACGAAATTAATTTCTAGATGATTAAGAAATGATTAATTAATTAATTAAT 4299
QY 4015 TATTAAGACCTGAGCTCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4074
DB 4300 TTTAAACATTAATCTCAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4359
QY 4075 GAGAAAGAGGATTAATTTCTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4134
DB 4360 GACAAAAAACACACTTTTAAAAAGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4419
QY 4135 ATCAATTTGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4194
DB 4420 ATTAATTTGACAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4479
QY 4195 GACCTTGACTTATTAATAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4230
DB 4480 GATTTAGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4539
QY 4231 -----CAAGCGAATGAATAATATCTG 4254
DB 4540 CCAACACACCTCCGTCACCGCAAAACAGACAGCAAAAGAGAAAGTAAGTTCTT 4599
QY 4255 CCCTTCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4314
DB 4600 CCATTTTAAACAACTGAGACCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4659
QY 4315 GTATTTGACCTGAGGAGCCAGCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4374
DB 4660 TTAATTTAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4719
QY 4375 AAATTAAGAGAGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4434
DB 4720 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4479
QY 4435 AATAACAATTTGCTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4494

OY	2776	CTGAAACTGTTTGAGAACATCTGTCTCTCGCAGAGATTAAGACAT---	CTACCAGA	2832
Dd	3217	TTGGAAGGATTGGTACTCTTAAATCTGGAAAAAATAAGTAGCTAATTCATTA	3276	
OY	2893	CTTATTGGACGAATATGTCGCGAAGCTCTGCAGAAATCTCAAGTCAAGACGACACA	2892	
Dd	3277	ACCATTTCTACACAGATGAGAAAATTTTATGAGAAATTTTTTAAAAAATATATATCC	3356	
OY	2893	TTCTATACGAGAGCTTCACTAACTCTGTGMAATCTAAAGCCGATATCAACTCTT	2952	
Dd	3337	TATTTTATGATGATATCAACAATTCGTAAAACTAATTCAAAGTTGATACAGGTTG	3396	
OY	2953	AACGATGAATCTAAAGCTTAAGAAAGCTGGAAGAGCATATAAGCTGAGACAGACTG	3012	
Dd	3397	ACCGAAACACAAAA---AATGCAATTAAATGAAATTTAAAAAATTAAAAAGATACCTTA	3453	
OY	3013	CAACTGAGCTTCGACGCTGCACACAGTCAAACTCAAACTGGAAGAGCTCTGCACAG	3072	
Dd	3454	CAGTATCATTTTGTATTATATTAATTAATTAATTTAAATTTAGATAGATTATTTAATAG	3513	
OY	3073	AAGAAGACAGTCGGCACAGTAAAGATGCGATCAAGTCAAGTCTGCTCAAGAGCAG	3122	
Dd	3514	AAAAAGAACTGGCCAAAGACAAATATGCCAATTTAAAAAATTTAATTTAAAGAACAA	3573	
OY	3133	CTTGAAAGCAAACTCAACTACTGTAACCAATCCGAACAGTACTGACAGACTTCTCACTG	3192	
Dd	3574	TTAGATCAAAATTAAGTAACTTCAATTAACCCACATTAATGTATTTCAAAACCTTTCTGT	3633	
OY	3193	TTCTTCAACAGAAGGAAGGAGCGAGTGCAGACAGACAGACACACTCTGGAAGAACCC	3252	
Dd	3634	TTCTTTAACAAAAAAGAAAGAGCTGAAATATGCAAGAACTGAAACACACTTTGAAAAACCA	3693	
OY	3253	AAGATTCTTCCAAACACTCAAAAGCCCTGTCAGATATTATTAATAGGAGAGTCTCTCT	3312	
Dd	3694	AAATATATTAATGAAACATTTATAAGACCTTTGTAATATTTAATATGATGAATCATCTCCA	3753	
OY	3313	CTGAAGACTCTCTCCGAGGAGACATTCAGACCCGAGGATTAACGCCACGCTGAGAAC	3372	
Dd	3754	TTTAAAACTTTAAAGTAAAGATCAATTAATTCAAACGAAAGATTAATTAAGCAATTTAGAAAAA	3813	
OY	3373	TTCAAGTCCGTGTAGCTGCAAGGCAAGCTBAAGCAACCTCAACCTGGAAGAAAG	3432	
Dd	3814	TTTAGAGTATTAAGTATAAATAGATGAAAACTCAATGATTAATTTACATTTAGAGAAAGAA	3873	
OY	3433	AAGCTCAGTACCTCTCTAGCGGACGTGCATCACCTGATCGCCGAGCTCAAGSAGTCAAT	3492	
Dd	3874	AAATTAATCTTTCTTAATCAAGTGGAATTAACATTAATTAAGTAATTAAGAAAGTATA	3933	
OY	3493	AAGAACAAGAACTACACCGCATTAAGCCACGAGAGATTAATACAGAGCTGATTAACCA	3552	
Dd	3934	AAAAATTAATAATTTACAGSTAAATCTCCCAAGTGAAAAATTAAGAAAGTTAAACGAACT	3993	
OY	3553	CTGGATCTTACAAAGAAGTTCGAGCTGGAAGGAACGATGTGC-----	3596	
Dd	3994	TTAAAAATCTTAACGAAATTTTCTCCAGAAAGCAAAAGTTAACACAGTTTAATCTCACCT	4053	
OY	3597	-----CACTGSGTGTCTGAATCTGTGGCTCCGACACTGAGCAAGCTTCA	3641	
Dd	4054	CAACCAATGTAACTCCATCTCCATTAATCTGTAGAGGTAAAGTAGTTACAGATCCACA	4113	
OY	3642	ACCTAAGAA-----GCCTGATCTACTCATGTGCGAGCCGAGTCCAAATTAACATTAAC	3693	
Dd	4114	AAAGAGAAACACAAATTAACCAACTTCAGGCTTTTATTAACGAATTTACAAACAGTACTA	4173	
OY	3694	ACATCTCAGAACTGCAGATGAGAGTGCATGACGTATCATTTGTGCTATCTTGGCGAG	3733	
Dd	4174	CAATTAACAAAATTTATGACGAAGAAGATGATCTCTTGTTGTATTTACCAATTTTGGAGAA	4233	
OY	3754	AGCGAGGAGCATCTGATGAGCTTCGCGCAGGAGGTGACCGGAGAGCTCACTCTTCC	3813	
Dd	4234	TCCGAGATTAATGCGAATTTTATGATCAATGATGTAAGTACGTGGAAGCAAT--ATCTGTC	4290	
OY	3814	GTAATGTAATCAATTCGTCCAAATATCGAAGCAATACGAAGTCACTGATCTGAACCT	3873	

[illegible]

[illegible]

OY	247	GGGTCGTGGCCCTCTGGGGGTTCCGTCGCCCTCCGGCGCAGCGTGGCATCAGTGGCTCA	306
Db	181	GGCTCAGTGGTCTTCAGGTGGCTCAGTTCGCTTCAGGTGGCTCAGTTCGCTCAGTGGCTCA	240
OY	307	GTGGCAAGCGCGGGTTCCGGGAACAGTCGAAGAACCAATCCATCTGCACACTAGCAT	366
Db	241	GTTCGCTTCAGGTGGTTAGGTAATTCACAGCGTCACAAATCTTCAGATPAATTCAGGTAT	300
OY	367	TCCGAGCGCAAGTCCTACGGCCGACCTCAGACCGAGTGAGAACTATCTCCTCAATC	426
Db	301	TCAGATCGTAATATCTACGTGATTTAAAAACAGAGTACGAAATTAATCTGTAACTATC	360
OY	427	AAGAGCGTAGTATCCCAACAGTTGTGCACCTCAGTCAATCATATGCTGCATGTGTAT	486
Db	361	AAAGACTCAAAATCTCTCACTCTTGATTTAGTAAATCATATGTATTAATCTTTGTGTAT	420
OY	487	AACATTCATGGCTTCAAAATCTGATTTGACGTTACGAGAGATCAATGACTCCTGTAC	546
Db	421	AATATTCATGGTTCAAAATTTTAATGATGATGATGAAGAAATTAATTAATTAATATAT	480
OY	547	AAATTGAATTTCTACTTCGACTTCGTATGAGGCCAAACTGAATGACGTTTGGCCCAATGC	606
Db	481	AAATTAACCTTTATTTTGTGATTTATTAAGAGCAAAATTAATGAATGTATGTCTATGAT	540
OY	607	TATTCGCAAAATTCATTCATTTTGAAGATCGAGCCACAGAGTTGGACGTATTTGAAGAG	666
Db	541	TATTCGCAAAATTCATTCATTCATTTTGAAGATCGAGCCACAGAGTTGGACGTATTTGAAGAG	600
OY	667	TTTGGCTTCGGATATCCGCAAGCCTCTCGACACATCATAGAGACAACTGGGAAAGATGGA	726
Db	601	CTTGTGTCGGATATTAAGAAAACCATTAAGCAAAATTTAAAGATATCTAGGAAAAATGGA	660
OY	727	GATTATATTTAAAAAGATTAGAAGACCATCGAAGACATTAACGAGCTGATTCGAAGATCC	786
Db	661	GATTATATTTAAAAAGATTAGAAGACCATCGAAGACATTAAGATTAATTAATTAAGAAAGT	720
OY	787	AAAAAGACCATAGCAAAAAATAGATTCACACCAAGAGAGAGAAAAAGAAAGTTGTAC	846
Db	721	AAGAAACAAATGTATTAATAATAGAAATGCAACTTAAGAAAGAAAGAAAAAAATTAATAC	780
OY	847	CAGGCGCCGTCGACCTGTCATCTATTAACAACACACTTGAAGAACCCATCACTCATC	906
Db	781	CAGCTCAATATGATCTTCTCATTTACAAATAAACATTTAGAAAGACACATATATTATA	840
OY	907	AGCGTACTGAGAAAGCCATAGACACCCTCAGAGAGATGAATAATATCAAGACTGCTC	966
Db	841	AGCGTTTATAGAAAAAGCTATTTAGCACTTTAAAAAAAATGAAACCTTAAGAAATTTACTT	900
OY	967	GACCAAGATTATGAATTTAAGAATCCCTCGCCGACCACTCTGGGAACCCCTTAACAG	1026
Db	901	GATTAAGATTAATGAATTTAATAATCCCCACCGCCCAATTTCTGGAATACACCAATACT	960
OY	1027	CTGCTGCAAGAAACAAGAGATAGAGAGCACAGAGAAAGATCAAAAGATCGCCAAA	1086
Db	961	CTCCTTGATAGAACAAAAAATGAGAGAACAGAAAAAGAAATTAAGAAATTTGCCAAA	1020
OY	1087	ACCATTAAGTTCACATATGATCTCTCTTTACTGATCCCTTGAGAGTGAATACATCTG	1146
Db	1021	ACTATTTAAATTTAATTTAGATGTTATTTTACTGATCCACTGTAATTTAAATATTAATTA	1080
OY	1147	AGAGAGAGATTAGAAATTTAGACATCTCGCCCAAGTGGAGACAGAGAAATCAACCCGA	1200
Db	1081	AGAGAAAAAATTAATAATATTGATATATAGTGCAGAAAGTTGAACAAAGAAATCAACTGA	1144
OY	1207	CCTAATGAAATATCCCAATGTTGACGTACCTCTCTTTATTAAGATATVCAACAACGT	1266
Db	1141	CCCAATGAAATATCCCAATGAGTACTTATCTTTCTCATATATACATATTTAAACAATGCT	1200
OY	1267	CTCAACGAGCTCATAGCTTCGCGTACCTTATTAACCCCTTGATATATACGAAAGAAACCC	1326
Db	1201	TTAATGAACCTTAATCTCTTTGGGATTTAATTAATTCATTTGATTAATTAACAAGAACCA	1260

QY	3788	TCACGGGTGAGGCGTGCACCTCTCCGTGATGATTAACATTTCTGTCACAAATCGGAAGC	384.47
Db	481	TACACGAGAGACGATCACTCCCTCCGTATTTGATTAACATCTTTCTAAATTTGAAATG	540
QY	3848	AATACGAAGTCTCTATCTGAAAACCTTGCGACGGCGCTATAGTCTCTCAAGAAACAGC	390.7
Db	541	AATATGAGGTTTTATATTTTAAACCTTTTGACGGCTGTTATGAAAGTTTAAAAACAAT	600
QY	3908	TGCGAATTAACGATGATGACCTTCATGTCACGTGAGGAGCACTTCGTAACGCGCCTTGA	396.67
Db	601	TAGAAATTAACGTTATATGACCTTTAATGTTTAAGGTTATTTTAAATTCACAGATTTA	660
QY	3968	ATAAGAGAGAAAATTTCCAGAACGCTTGAGAGCGCACTTGATTCCTATAAAGACTGA	402.77
Db	661	ATAAAGCTGAAAATTTCCAAAATGTTTTGAAATCAGATTTATTCATATTAAGATTTAA	720
QY	4028	CCCTCCTTACTACGTGTTGTGAAGACCCATACAGATTCCTCTCAATTAAGAGAGGGATA	408.7
Db	721	CATCAAGATTAATATGTTGTCTTAAAGATTCATATAATTTGTTATTAAGAAAAAAGATTA	780
QY	4088	AATTTCTGTAGTTTACCACTATATCAGAGACTCCATCCACACCATATCATATTTCCGTA	414.7
Db	781	AATTTCTTAAGACAGTAAATATATTAATTAAGGATTCATATAGCGATTAATAATTTTGAA	840
QY	4148	ATGATGTCTGGGTTATTAACAATCTCTGAGCAAAAATACAGTCTGACTGACTTA	420.7
Db	841	ATGATGTCTTGGATTAATTAATAATTTATCCGAAAAATATAATCAGATTTAGATTTAA	900
QY	4208	TTAAAAAGTATATACACGATATACAGAGCGAGATGAAAATATGTGCTTCCGATTA	426.7
Db	901	TTAAAAATATATCAACAGACAAACAGTGAAATAGGAAATACCTCCCTTTTAAACA	960
QY	4268	ACATCGAAACCCCTGTACAGACAGTGAACAGACAAAATGCACTCTTTCGATTAACCTGG	432.7
Db	961	ATATATTGACCTTATATAAAAAAGTTAATGATTAATTTATTTATTTGTAATTTCAATTAG	1020
QY	4328	AGGCCAAGTCCCTCACTATATACGTACGAGAGGCAATGTGGAAGTTAAATCAAGAGAC	438.7
Db	1021	AAGCAAAAGTTCTAAATTTATACATATGAGAAATCAACGCTAGAAAGTTAAATTAAGAAC	1080
QY	4388	TGAACCTACTCAAAACAATCCAAAGACAGCTGGCAGATTTCAAGAAAAATAACAAATTCG	444.7
Db	1081	TTAATTTACTTAAAAACAATTCACAGACAAATGTGCAATTTTAAAAAAAATAACATTTTCG	1140
QY	4448	TGCGAATTTGACAGCCCTGTACGATTTATTAACCAACAAGATTCCTGACCAAGTTTCTGT	450.7
Db	1141	TTGGAATTTGCGATTTTATTCACAGATTTATACCATATATACTTATTTGACAAAGTTCTTA	1200
QY	4508	CCACTGCGCATGTGTTTGCAAAACCTCGCCAAAACAGTGTGAGCAATCTGCTCGACGGCA	456.7
Db	1201	GTACAGGATATGTTTGTGAAAATCTTGCTATAAACGTTTTATCTATATTTACTGTATGAA	1260
QY	4568	ACCTGCGACGGGATGCTGACATCTCCAGACCAATGCTGTAAGAAACAGTGCCTCCGCA	462.7
Db	1261	ACTTGCAGAGTATGTTTAAACATTTTCAACACCCAAAGCCTTAAAAAACAATGTCCACAAA	1320
QY	4628	ATACGCGCTGTTTACGACCTGTGAGAGAGGGCAAGATGACAGTGTCTCCGTAACATCA	468.7
Db	1321	ATTTCTGATTTTTCAGACATTTTGAATGTAAGAGAGACAGATGTAAATGTTTATTAATTTCA	1380
QY	4688	AACAAAGAGAGATTAAGTGTGTGGAGAACCCCAAAACCTTACCTGCAATGAAAACAATGCGC	474.7
Db	1381	AACAAAGAGAGATTAAGTGTGTGGAGAAATCCAAATCTCTACTGTATACGAAATATATGTG	1440
QY	4748	GGTGTACGCCGATGCTAAATGCAACCGAGAGACACGGCGCTCTAACGGAAGAAATATCA	480.7
Db	1441	GATGTGTATGACAGATGCCAAATGTACGGAAGAGATTCAGGTGACCAACGGAAAAAATCA	1500
QY	4808	CATGCGAGTGTACTTAACCCGAGCTCCTATACACTCTGTGACGGGATTTTTTTTCTCCACT	486.7
Db	1501	CATGTATGTACTTAACCTGATTTCTTATCCACTTTTTCAGATGATATTTTCTGCACTTCTC	1560

QY	4868	CTAATTTCCTGGGATCCGTCCTCCGCGTGAACCTATCTGTCGTCACAGCTCACT	4927
Db	1561	CTAACTTCTTGAAGATATCATCTTATTAAATACATCATGTATAATATACAGTTTCATT	1620
QY	4928	AATAGATGATG	4939
Db	1621	AAAAATGTAGG	1632
RESULT 19			
LOCUS	PFAMSP808		
DEFINITION	PFAMSP808	1636 bp	DNA
ACCESSION			linear
VERSION			INV 29-MAY-2002
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
SOURCE			
gene			
CDS			
BASE COUNT			
ORIGIN			
Query Match	18.4%	Score 910.4:	DB 3; Length 1636;
Best Local Similarity	72.4%;	Pred. No. 3.3e-189;	

Matches 1181; Conservative 0; Mismatches 451; Indels 0; Gaps 0;

```

OY 3308 CCCCCTGAAGAGCTCTCCGAGAGAGATCCAGACCAGATTAACCTGCGCAGCTCG 3367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CCCCCTGAAGAGCTCTCCGAGAGATCCAGACCAGATTAACCTGCGCAGCTCG 60
OY 3368 AGAATCTCAAGCTCTGCTGTAAGCTGCAAGGACGACGACGACGACGACGACG 3427
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 AAAACTTAAAGTATTAAGTAAATTAAGAGAAATTAAGATTAATTAATTAAGAAA 120
OY 3428 AGAAGAGCTCAGTACCTCTCTAGCGGACGACGACGACGACGACGACGACG 3487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AGAAAAATTAATCACTACTATCAAGGATACATCATTTAATGCGTAATTAAGAA 180
OY 3488 TCATTAAGACAAAGATCAACCGGCAATAGCCGACGACGACGACGACGACG 3547
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TAAATAAAAATTAATTAAGGTAATTCACAAAGTCAAAATTAATAGGATGTTAA 240
OY 3548 ACCGACTGGAATCTTACAAAGAGTCTGCTGGAAGGAAAGATGCGCAGCTGCTG 3607
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 ATGCAATTAAGATCTTACAAAAATTTCTCCAGAAAGAAAGATGTTGCAACAGTT 300
OY 3608 CTGAATCTGCTCCGACACACTGAGACAGTCTCAACTAAGAGCTGCACTACTG 3667
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GTGAAGTGGATCCGACACATTAAGAAAGTCAACCAAGAAACGACATCACTCATG 360
OY 3668 TCGGAGCCGAGTCAATCAATTAACCACTCAAGACGACGACGACGACGACG 3727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 TAGGAGCAGAGTCTAACCAATTAACCACTCAAGAGTCTGATGATGAGTATGAG 420
OY 3728 TCATTAATGCTCTCTCTGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 3787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TAATCATAGTACTTATTTTGGAGATCCGAAAGAAATTAATTAATTAAGCAAG 480
OY 3788 TCACCGTGAAGCTCTCACTCTCTGATGATTAATCACTCTGCTCAAAATCG 3847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TAACGAGGAAAGCAGTAACTCTCTGATTAATCACTCTCTCAAAATCGAAATG 540
OY 3848 AATAGAGTGTCTATCTGAAAGCTGCGAGGAGGAGGAGGAGGAGGAGGAGG 3907
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AATATAGAGTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600
OY 3908 TGGAGATTAAGCTGATCACTTCAATGTCACGAGGACATCTGACAGCCCTT 3967
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 TAGAAATTAAGCTGATGACATTAATTAATTAATTAATTAATTAATTAATTA 660
OY 3968 ATAAGAGAAATTAATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4027
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 ATAAGAGAAATTAATTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
OY 4028 CCTCCTTAAGTGTGCAAGAGCCATACAGTCTCTCAATTAAGAGAGGAGG 4087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CATCAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
OY 4088 AATTTCTGTAGTCAACTATTAATCAAGGACGACGACGACGACGACGACG 4147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 AATCTTAAGAGTATTAATTAATTAAGATTAATTAATTAATTAATTAATTA 840
OY 4148 ATGATGTCTGTGATTAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 4207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 ATGATGTCTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
OY 4208 TTAATAAGATTAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 TTAATAAGATTAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
OY 4268 ACATGAAACCTGTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 AATATGAGACCTTATTAATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
OY 4328 AGCCCAAGTCTCAACTATTAATTAAGAGGAGGAGGAGGAGGAGGAGGAGG 4387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 AAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1080

```

```

OY 4388 TGAACCTCAAAACCAATCAAGACAGCTGGAGATTTCAAGAAAAATTAACATTTG 4447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
OY 4448 TCGGAATTCAGACCTGCTTACCGATTAATTAACCAACATCTCTGACCAAGTTCTG 4507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 TTGGAATTCGATTTATTAACAGATTAATTAACCAATTAATTAATTAATTAAT 1200
OY 4508 CCACCTGAGGCTGTGTTGCAAAACCTGCAAAACAGTGTGAGCAATGCTGACAGG 4567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 GTACAGGATGATGTTTGAATAATCTTGCAAAACGTTTATCTAATTAATTAAT 1260
OY 4568 ACCTGAGGAGGATGCTGCAACATCTCCAGCACCAGTGCCTGAAGAAACAGTCC 4627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 ACTTGCAAGGTATGTTAAACATTTTCAACACCAATGCTTAATAAATGTCACAA 1320
OY 4628 ATAGCGGCTGTTTCAAGGCTGAGAGAGCGGAGAGTCAAGTCTCTGCAACTCA 4687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1321 ATTCGTGAGTGTTCAGACATTTAGATGAAGAAAGAAATGTAATTTATTAATTA 1380
OY 4688 AACAGAGAGATTAAGTCTGAGAGAACCAACCTTACCTGCAATGAATAATGCGG 4747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1381 AACAGAGAGTATTAATGTTGTTGAAATCCAAATCTCTGTAACGAAATTAATG 1440
OY 4748 GGTGAGAGCCGATGCTAAATGACACCGAGAGAGAGAGAGAGAGAGAGAG 4807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1441 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
OY 4808 CATGCGAGTGTACTTAAGCCGACCTCTATCCACTCTGAGCGGATTTTCTCTCAG 4867
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 CATGGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
OY 4868 CTAAATTCCTGGGATCTCTCTCTGCTGATGATGATGATGATGATGATGAT 4927
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1561 CTAAATTCCTAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
OY 4928 AATAGATGATG 4939
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1621 AAAAAATGATG 1632

```

RESULT 20
PFAMSP828 1636 bp DNA linear INV 29-MAY-2002
LOCUS
DEFINITION
P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.
D13360
D13360.1 GI:391798
ACCESSION
EGR-like domains; major merozoite surface protein precursor.
KEYWORDS
Plasmodium falciparum (isolate 828) merozoite, DNA.
SOURCE
ORGANISM
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 1636)
Jongwutives, S., Tanabe, K. and Kanbara, H.
Amino acid sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
93295445
MEDLINE
Jongwutives, S.
Jongwutives, S.
AUTHORS
Jongwutives, S.
TITLE
Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex. 3747), Fax:0958-47-6607)
COMMENT
Submitted (05-OCT-1992) to DDBJ by:
Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi

RESULT 21
PFAMSP837 1636 bp DNA linear INV 29-MAY-2002
LOCUS P. falciparum DNA for the precursor to the major merozoite surface
DEFINITION proteins, C-terminal.
ACCESSION D13362.1 GI:391806
VERSION D13362.1
KEYWORDS Egr-Like domains; major merozoite surface protein precursor.
SOURCE Plasmodium falciparum (isolate 837) merozoite, DNA.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.
REFERENCE 1 (bases 1 to 1636)
AUTHORS Jongwutives, S., Tanabe, K. and Kanbara, H.
TITLE Sequence conservation in the C-terminal part of the precursor to
the major merozoite surface proteins (MSP1) of Plasmodium
falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
REFERENCE 2 (bases 1 to 1636)
AUTHORS Jongwutives, S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical
Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4
Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747),
Fax:0958-47-6607)
COMMENT Submitted (05-OCT-1992) to DDBJ by:
Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source Location/Qualifiers
1. 1636
/organism="Plasmodium falciparum"
/db_xref="taxon:3833"
/chromosome="9"
/dev_stage="merozoite"
1. 1622
/gene="MSP1"
<1. 1622
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BAA02623.1"
/db_xref="GI:391807"
translation="PLKLTSESIOTEDNNTASLENFKVLSKLEGLKLNLEKKLS
YLSGLHLIALEKLVINKNTYGPSVNNITDVNNALLESYKFLPESTDVATVSES
GSDLEDSQPKPASTHVGAESNITTSQNVDDVDVITVIFESSEDDLDLQVY
TGEAVTPSVINILSIENEYELVLEKPLAGYRSLEKOLENNVTFVNVYDILNSR
ENKRENFNVLESDLIPYKDLTSSNVVKKDYKFLNKKRDKFLSSVYINDSIDTL
NEANDVGYRIILSEKYSKSDLSIKYINDQGEKEKLPILNIEITIKYVNDKIDL
FVHLEKAVLNTTTERKSNVEYKIKELNLTKTIDKLAQFKKNNNVGTADISDHN
NLTKFLSTGAVFENLAKTVLSNLDNMLQGLNLSOHQVKKKQCPQNSGCFRIHDER
EECKCLNFKQEDKCVENPPTCNENNGGDADDAKCTEEDSGSNGKITCECTPDPS
YPLFDGICSSSNFLGISFLILMLILYSFI"
BASE COUNT 694 a 199 c 239 g 504 t
ORIGIN
Query Match 18.4%; Score 908.8; DB 3; Length 1636;
Best Local Similarity 72.3%; Pred. No. 7.5e-189;
Matches 1180; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

QY 3308 CTCCTCTGAGAGCTCTCTCCGAGAGAGATCCAGACCGAGATACCTACGCCAGCTCG 3367
DB 1 CTCCTTTAAAACTTAAGCTTAAGTAATTCAAACAGAAATATATATGCGCAGTTTG 60
QY 3368 AGAATTCAGAGTCTGTCTAGCTGCAAGCAAGCTGAAGCAACCTGAACCTTGAGA 3427
DB 61 AAAACTTTAAAGTATTAAAGTAAATTAAGCAAAATTAAGATTAATTTAAATTTGAAA 120
QY 3428 AGAAGAGCTCAGTACCTCTCTAGCGAGCTGATCACCCTGATCGCGAGCTCAAGAG 3487
DB 121 AGAAAAATTAATCACTTATCAAGTGAATTCATCATTTAATTCGTGAATTAAGAG 180
QY 3488 TCATTAGAACAAAGAACTACCGGCAATAGCCCAAGCAGATATACAGAGCTAATA 3547
DB 181 TATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
QY 3548 ACGCACTGAGATCTTCAAGAAAGTTCTCGCTGAAGAGCAAGATGCGCAGCTGTGT 3607
DB 241 ATGCATTAAATCTTCAAAAAATTTCTCCAGAGAGCAAGATGTTGCAACAGTTGTA 300
QY 3608 CTGAATCTGGCTCCGACACAGCTGAGACAGTCTCAACCTTGAAGAGCTGATCTATG 3667
DB 301 GTGAAGTGGATCCGACACATTAGAACAAAGTCAACCAAGAAACCAAGCATCACTG 360
QY 3668 TCGAGCCGAGTCCATACATTAATACCATCTCAGACCTCGACATGAGTCTGATGAG 3727
DB 361 TAGGACGACAGCTTACACAAATTAACAAATCAACAAATGTCAGATGATGATGAGAG 420
QY 3728 TCATCTATGCTGCTATCTTCGCGAGAGAGAGAGTACGATGACTCGCGCAGCTG 3787
DB 421 TATCTATGCTATCTATTTGGAGATCCGAGAGATATGATGATTTAGGCAAGTGA 480
QY 3788 TCACCGTGAAGCTGCTACTCTCCGATGATTAACATCTCTGCAAAATCGAGACG 3847
DB 481 TAACAGAGAGAGAGTAACTCTCCGATGATTAACATCTCTGCAAAATCGAGACG 3907
QY 3848 AATGCAAGTGCCTCTCTGCAAACTCTGCGAGGCTGATGATGATGATGATGATGAT 600
DB 541 AATGAGGTTTATATTTAAACCTTGAACGCTTATGAGAGTTTAAATTAAGAAATG 600
QY 3908 TGGAGATACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3967
DB 601 TAGAAATTAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 3968 ATTAGAGAGAAATTTCAAGAGAGCTTGGAGAGCAGCTTGAATTCCTTAAAGACTGA 4027
DB 661 AATAAGCTTAAATTTCAAAATGTTTGAATTAATTAATTAATTAATTAATTAATTA 720
QY 4028 CCTCTCTACTAGCTGCTGCAAGAGCCTATACAGTATGATTAAGAGAGAGATA 4087
DB 721 CATCAAGTATTAATGTTGCAAGATCCATTAATTTCTTAATTAAGAAAAAGAGATA 780
QY 4088 AATTTCTGTAGTTACAACTATATCAAGAGCTCATGACCCGATATCAATTTGCTTA 4147
DB 781 AATTTAGAGAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 840
QY 4148 ATGATGCTGTGGGATATTAACAGATCTGAGGAGAAATTAACAGTGTGACTGCTTA 4207
DB 841 ATGATGCTGTGGGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
QY 4208 TTAAGAAATATCAAGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4267
DB 901 TTAAGAAATATCAAGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 4268 ACATCAAGACCTGTGCAAG 4327
DB 961 ATATTGAGAGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1020
QY 4328 AGGCAAGTCTCTCAACTTACTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4387
DB 1021 AAGCAAGTCTCTCAACTTACTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 4388 TGAATCACTCAAAACAAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4447
DB 1081 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140
QY 4448 TCGAATGTCAGAGCTGTCTACGATTAATTAATTAATTAATTAATTAATTAATTAAT 4507

Db 1141 TTGGAAATGCTGATTTATATCAACAGATTATTAACCTTAATCTATTGACAAAGTTCTTA 1200
 QY 4508 CCACTGCGATGCTGTTCGAAAACCTCGCAAAACAGTGTGAGCAATCTGCTCGACGCA 4567
 Db 1201 GTACAGGTATGCTTTTGAATAATCTTGTCAAAACCGTTTATCTAATTTACTTGATGAA 1260
 QY 4568 ACTGCGAGGCGATCTGACATCTCCAGACCAACGATGCGTGAAGAAAGTGCCTCAGA 4627
 Db 1261 ACTTGAAGGTATGTTAAACATTTTCAACACCAACGATGCTAAAAAACAATGTCCACAA 1320
 QY 4628 ATAGCGGCTTTTCAGCACTGAGAGGCGAGAGTGCATGCTCTCTGCACTACA 4667
 Db 1321 ATCTGATGTTTCAGCATTTTATGATGATAAGAGAAATGTAATGTTTATTAATTTACA 1380
 QY 4688 AACAAAGAGAGATATAGCTGTGAGAGAACCAACCTTACTGCAATGAATAACATGGCG 4747
 Db 1381 AACAAAGAGGTATTAATGTTGTAATAATCCAAATCTTACTTGTAAAGAAAATATGCTG 1440
 QY 4748 GGTGTGACGCCGATGCTAAATGACACCGAGAGAGACAGCGCTCTAACGAAAGAAATCA 4807
 Db 1441 GATGTGATGAGATGCAAAATGTACGAGAAAGATTCAGTAGCAGCAAGGAAAGAAATCA 1500
 QY 4808 CATGCGAGTGTACTAACCCGACCTCTTATCCACTCTTCGACAGGATTTTTCCTCCAGCT 4867
 Db 1501 CATGTGATGATCTAAACCTGATTTCTTATCCACTTTTCGATGATTTTCTGCAAGTTCT 1560
 QY 4868 CTAAATTTCTGCGCATCTCTCTGCTGATCCCTCATGCTGATGCTGATGCTGATGCTTCACT 4927
 Db 1561 CTAACTTCTTATGAGATATCTTATCTTATTAATCTGATGATGATTTATATATATACAGTTTCA 1620
 QY 4928 AATGATCGATG 4939
 Db 1621 AAAAATGTAGG 1632

RESULT 22
 PFAMSP79 1636 bp DNA linear INV 29-MAY-2002
 LOCUS P. falciparum DNA for the precursor to the major merozoite surface
 DEFINITION proteins, C-terminal.
 ACCESSION D13356.1 GI:391824
 KEYWORDS BGF-like domains; major merozoite surface protein precursor.
 SOURCE Plasmodium falciparum (isolate T9/94) merozoite, DNA.
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 REFERENCE 1 (bases 1 to 1636)
 AUTHORS Jongwutives,S., Tanabe,K. and Kanbara,H.
 TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
 JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
 MEDLINE 93295445
 REFERENCE 2 (bases 1 to 1636)
 AUTHORS Jongwutives,S.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology: 12-4 Sakamoto-machi, Nagasaki 852, Japan (tel:0958-47-2111(ex.3747), fax:0958-47-6607)
 COMMENT Submitted (05-OCT-1992) to DDBJ by:
 Somchai Jongwutives
 Department of Protozoology
 Institute of Tropical Medicine
 Nagasaki University
 12-4 Sakamoto-machi
 Nagasaki 852
 Japan
 Phone: 0958-47-2111x3747
 Fax: 0958-47-6607.
 Location/Qualifiers
 1..1636 /organism="Plasmodium falciparum"

Query Match 18.4%; Score 908.8; DB 3; Length 1636;
 Best Local Similarity 72.3%; Pred. No. 7,5e-189;
 Matches 1180; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

BASE COUNT 695 a 199 c 239 g 503 t

ORIGIN

Db 1141 TTGGAAATGCTGATTTATATCAACAGATTATTAACCTTAATCTATTGACAAAGTTCTTA 1200
 QY 4508 CCACTGCGATGCTGTTCGAAAACCTCGCAAAACAGTGTGAGCAATCTGCTCGACGCA 4567
 Db 1201 GTACAGGTATGCTTTTGAATAATCTTGTCAAAACCGTTTATCTAATTTACTTGATGAA 1260
 QY 4568 ACTGCGAGGCGATCTGACATCTCCAGACCAACGATGCGTGAAGAAAGTGCCTCAGA 4627
 Db 1261 ACTTGAAGGTATGTTAAACATTTTCAACACCAACGATGCTAAAAAACAATGTCCACAA 1320
 QY 4628 ATAGCGGCTTTTCAGCACTGAGAGGCGAGAGTGCATGCTCTCTGCACTACA 4667
 Db 1321 ATCTGATGTTTCAGCATTTTATGATGATAAGAGAAATGTAATGTTTATTAATTTACA 1380
 QY 4688 AACAAAGAGAGATATAGCTGTGAGAGAACCAACCTTACTGCAATGAATAACATGGCG 4747
 Db 1381 AACAAAGAGGTATTAATGTTGTAATAATCCAAATCTTACTTGTAAAGAAAATATGCTG 1440
 QY 4748 GGTGTGACGCCGATGCTAAATGACACCGAGAGAGACAGCGCTCTAACGAAAGAAATCA 4807
 Db 1441 GATGTGATGAGATGCAAAATGTACGAGAAAGATTCAGTAGCAGCAAGGAAAGAAATCA 1500
 QY 4808 CATGCGAGTGTACTAACCCGACCTCTTATCCACTCTTCGACAGGATTTTTCCTCCAGCT 4867
 Db 1501 CATGTGATGATCTAAACCTGATTTCTTATCCACTTTTCGATGATTTTCTGCAAGTTCT 1560
 QY 4868 CTAAATTTCTGCGCATCTCTCTGCTGATCCCTCATGCTGATGCTGATGCTGATGCTTCACT 4927
 Db 1561 CTAACTTCTTATGAGATATCTTATCTTATTAATCTGATGATGATTTATATATATACAGTTTCA 1620
 QY 4928 AATGATCGATG 4939
 Db 1621 AAAAATGTAGG 1632

Query 3308 CTCTCTGAGAGCTCTCTCCGAGAGAGCATCCAGACCGAGGATTAATCTAGCCAGCTCG 3367
 Db 1 CTCTCATTAATAAATCTTAAGTGAAGATCAATTCACAGAGATTAATTAATGCAAGTTAG 60

Query 3368 AGAATCTCAAGTCTCTGCTGTAAGTCAAGGCAAGCTGAAGACAACTGAACCTGAGGA 3427
 Db 61 AAAATCTTAAGTATTAAGTAAATTAAGAGAAATTAAGAGAAATTAATTAATTAAGAAA 120

Query 3428 AGAAGAGCTGACCTCTCTAGCGGACTGCATCCTGATGCGCGAGCTCAAGAGAG 3487
 Db 121 AGAAAAATTAATGATCACTTATCAAGAGATTAATCACTTAATTAATTAATTAAGAAAG 180

Query 3488 TCATTAAGAACAGAACTATACACCGGCAATAGCCACGAGAAATTAATACAGCTGATA 3547
 Db 181 TAATAAAAAATTAATAATTAATACAGTATTTCTCAAGTGAATTAATACAGTATTAATTA 240

Query 3548 ACGCATGGAATCTTACAAAGAGTTCTGCTGAGAGAGAGATGCGCAGCTGCTGCT 3607
 Db 241 ATGATTAAGATCTTACAAAAAATTTCTCCAGAGAGAGAGATGCTGCAACAGTTTAA 300

Query 3608 CTGATCTGGCTCCGACACACTGAGAGAGCTGCAACTGAAGAGCCCTGATCACTCATG 3667
 Db 301 GTGAAGTGGATCCGACACTTAAGAACAAAGTCAACCAAGAAACCAAGCATCAACTCATG 360

Query 3668 TCGAGCGGAGTCCAAATACATTAATTCACATCTCGAAGCTGAGAGAGTGCATGAGC 3727
 Db 361 TAGGAGCAGAGCTCTAACACATTAACACATACACAAATGTCAGATGATGATGATGATG 420

Query 3728 TCATCATTTGCTCTATCTTGGCGAGAGAGAGAGACTACGATGACCTGCGCCAGTGG 3787
 Db 421 TAATCATATGACCTATATTTGGAGAAATCCGAAGAGATTTGATGATGATTTAGCAAAATG 480

Query 3788 TCACCGGTGAGGCTGCTACCTCTTCCGTGATGATTAACATTTCTCCAAATAGAGAG 3847
 Db 481 TAACAGGAGAGAGCTAATCTCTCCCTTAATTAACATTAATTAATTAATTAATTAATTA 540

Query 3848 AATACGAGAGCTCTATCTGAACCTTGGAGGCGCTATAGGCTCTCAAGAAACAGC 3907
 Db 541 AATATGAGGTTTATTAATTAACCTTTAGAGGTTTATAGAGTTTAAAAAACAAT 600

Query 3908 TGGAGATTAACGATGACCTCAATGTCACAGGAGAGACATCTGACAGCGCTGTTA 3967
 Db 601 TAGAAATTAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 660

Query 3968 ATAAAGAGAAATTTCAAGAGCTTTGGAGAGCAGCTGATTCCTTAATTAAGACCTGA 4027


```

Db 661 ATAAACGGAATAATTTCAAAATGTTTAGATCAGATTATTCATATATAAGATTAA 720
Qy 4028 CCTCTTAAGTACTAGTGTGCAAGGACCATACAGTCTCTCATATAAGAGAGGATA 4087
Db 721 CATCAAGTAATTAATGTTCTCAAGATCATATAAATTTCTTAATAAAGAAAAAGATA 780
Qy 4088 AATTCTCTAGTATACAAATATATCAAGAGCTCCATGACACCATATTCATTTCCGTA 4147
Db 781 AATTCTTAAGCAGTTATATATATTAAGATTCAATAGATACGATATTAATTTGGCA 840
Qy 4148 ATGATGCTCTGGGTATTTACAAAGATCTGAGGAAAAATACAACTCTACCTTGCTTA 4207
Db 841 ATGATGTTCTGGATATATATTAATATATATCCGAAAAATATTAATACAGATTAGTCAA 900
Qy 4208 TTAATAATATATATCAACATTAAGCAGGAGATGAAAAATATATGTCCTTCTGTA 4267
Db 901 TTAATAAATATATCAACAAACAAAGGTGAATAAATACCTCTCCCTTTTAAACA 960
Qy 4268 ACATCGAAGACCTGTACAAAGACAGTGAAGACAAATGACCTCTTCTTAATTCACCTG 4327
Db 961 ATATTGACACCTTATATATAAACAGTAAATGATTAATGATTTATTGTAATTCATTAG 1020
Qy 4328 AGCCCAAGGTCCTCACTATATCTTACGAGAGAGCAATGTGGAAGTTAAATCAAGAGC 4387
Db 1021 AAGCAAAAGTTCTAAATATATACATATGAGAAATCAACGTAGAAAGTTAAATAAAGAAC 1080
Qy 4388 TGAATCTACTCAAAACATCCAGACAGCAAGCTGGCAGATTTCAGAAAAATATCAATTGCG 4447
Db 1081 TTAATCTACTTAAACAAATTCAGACAAATGCGAGATTTTAAATAAATAAACAATTCG 1140
Qy 4448 TCGAATTTGACAGACCTGTCTACCGATTTAAACACACAAATCTCTGACCAAGTTCTGT 4507
Db 1141 TTGGATTTGCGATTTATCAACAGATTTATACCATTAATACATTATTCAGAAAGTTCTTA 1200
Qy 4508 CCACTGCGATGTTGCGAAAACTCGCCAAACAGTCTGAGCAATGCTCGAGCGCA 4567
Db 1201 GTACAGGATGTTGTTTAAAAATCTTCTAAACCGTTTATCTAATTTACTTGATGGAA 1260
Qy 4568 ACCTGACGGGATGCTGAACATCTCCGACACCAATGCTGTAAGAAACAGTCCCGCCAGA 4627
Db 1261 ACTTGCAAGGATGTTTAAACATTTTCACACACCAATGCGTAAAAACAAATGTCACAAA 1320
Qy 4628 ATAGCGCTGTTTTCAGCATCTGACGAGCGGAGAGTGAAGTGTCTCTGACATACA 4687
Db 1321 ATTCGATGTTTTCAGCATTTAGATGAAGAGAAAGATTAATGTTATTAATTTACA 1380
Qy 4688 AACCAAGAGAGATATAGTCGTGGAAGAACCAACCTTACTGCAATGAAGAAATGCGC 4747
Db 1381 AACCAAGAGGTGATTAATGTTGTAATTCCAAAATCTTGTAAACGAAAAATATGCTG 1440
Qy 4748 GGTGTGAGCGGATGCTAAATGCAACGAGAAACAGGCGCTTAACGAAAAAATACA 4807
Db 1441 GATGGAATGACAGATGCCAAATGACGAAAGAAATTCAGATGACGAAAGAAAAATTA 1500
Qy 4808 CATGCGATGTAATTAAGCCGACTCTTACATCTTCTGACGAGATTTTTCCTCCAGCT 4867
Db 1501 CATGGAATGATTAATTAACCTGATTTCTTCCACTTTTGAATGATTTTTCGCACTTCT 1560
Qy 4868 CTAATTTCTGGGCAATCTCTTCTCTGATGATTCATGCTGATCTGATGACGCTTCACT 4927
Db 1561 CTAATCTTGAATATATCATCTTAATTAATATACATGATTAATTAATACAGTTTCATTT 1620
Qy 4928 AATAGATGATG 4939
Db 1621 AAAAAATGTAGG 1632

```

RESULT 23
 PFAMSP834
 LOCUS PFAMSP834 1636 bp DNA linear INV 29-MAY-2002
 DEFINITION P. falciparum DNA for the precursor to the major merozoite surface proteins, C-terminal.

```

ACCESSION D13361
VERSION D13361.1 GI:391800
KEYWORDS EGF-like domains; major merozoite surface protein precursor.
SOURCE Plasmodium falciparum (isolate 834) merozoite, DNA.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 1636)
TITLE Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 59 (1), 95-100 (1993)
MEDLINE 93295445
REFERENCE 2 (bases 1 to 1636)
AUTHORS Jongwutives, S.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-1992) Somchai Jongwutives, Institute of Tropical Medicine, Nagasaki Univ., Dept. of Protozoology; 12-4 Sakamoto-machi, Nagasaki 852, Japan (Tel:0958-47-2111(ex.3747), Fax:0958-47-6607)
COMMENT Submitted (05-OCT-1992) to DDBJ by: Somchai Jongwutives
Department of Protozoology
Institute of Tropical Medicine
Nagasaki University
12-4 Sakamoto-machi
Nagasaki 852
Japan
Phone: 0958-47-2111x3747
Fax: 0958-47-6607.
FEATURES
source
location/Qualifiers
1..1636
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="9"
/dev_stage="merozoite"
1..1622
/gene="MSP1"
<1..1622
/gene="MSP1"
/codon_start=3
/product="major merozoite surface protein precursor"
/protein_id="BAA02622.1"
/db_xref="GI:391801"
/translation="PLKLTSEESFOTEDNYASLEFVLSKLEKTLKONLLEKKLST
YLSGHLHLIAELKEVTKNNKNTGSPNNQVNNALSEYKKLTPGCTGATATPVSES
GSDTLEOSPQRKRPASTHVGAESNTITTSQVDDDEVYITVPIGESEEDYDLAGOVY
TGEAVTSPVIDNLSKLTENEVEVLYLPLAGVYRSKLEKLENNVMTNVNKKDLNLR
FNKRENFKNVLESDILPYKDLTSSNVVKKPKYKLNKLEKDKFLSSYNYTKDSIDTDI
NFANDVIGYKILSEKYKSDLSIKYINDKQENERYLPLNNIETLYRTVNDKIDL
FVHLLEKVLNVTYETKSNVEKIKELNLYKTIOKLDAPFKNNNFVGIADLSTDYNN
NLTKRFLSTGVNPENLAKYTLNLDNLGNLQGMNLSOHOCVKKQCPONSQCFRLDPR
BECKCLNTKQESKCYENPPTCNENNNGGCDADAKCTEEDSGSNKKITCECTKPS
YPLFDGIFGSSSNFLGISFLILMLILYSF1"
BASE COUNT 695 a 199 c 239 g 503 t
ORIGIN
Query Match 18.48; Score 907.2; DB 3; Length 1636;
Best Local Similarity 72.28; Pred. No. 1.7e-188;
Matches 1179; Conservative 0; Mismatches 453; Indels 0; Gaps 0;
Qy 3308 CTCCCTGAGACACCTCCGAGGAGACATCCAGAGAGATTAACGACGCTCG 3367
Db 1 CTCCATTAATAAATTAAAGTGAAGATCAATTTCAACAGAAAGATTAATATGCCAGTTAG 60
Qy 3368 AGAAGCTCAAGTCTGTCTAAGCTCGAAGGCAAGCTGAAGACAACTGAACCTGAGGA 3427
Db 61 AAAAGCTTAAGTATTAAGTAAATTAAGAGAAATTAAGAGATTAATTAATTAAGAA 120
Qy 3428 AGAAGACCTGACGCTCTAGCGACATCGATCATGCTGAGCGGAGCTCAAGGAG 3487
Db 121 AGAAAAAATTAATCAATCACTTAACAGTGAATTAATTAATTAATTAAGAGAG 180

```


QY	3488	TCCTTAAAGAACAAAGACTTACACCGGCACTGGCCAAAGCAAGTAATATACAGCCGTAAT	3547
Db	181	TAATTAATAAATAAATAATATATACAGTAATCTTCCAAAGTAATAATATACGATGTGTAACA	240
QY	3548	ACGCAGCTGGAAATCTTTACAGAAGTTCCTGGCTCAAGGAACAGATGTGGCACTGTGGTGT	3607
Db	241	ATGCATTTAGAAATCTTACAAAAAATTTCTCCAGAAAGAAACAGATGTGGCAAGTGTGTA	300
QY	3608	CTGAATCTGGCTCCGACACACTGGAGACGTCTAACCTAAGAAGCTGCATCTACTCATG	3667
Db	301	GTGAAGAGTGATTCGACACATTTAGAAACAAAGTCAACCCAAAGAAACCCAGCATCACTCATG	360
QY	3668	TGCGAGCCGAAATCCAAATACATTACCAATCTCAGAACCTCGACATGTGGTGGTGTGACG	3727
Db	361	TAGAGAGCAGATCTTACACAAATAAACAACATCAACAAATGTCCATGATGAAAGTATGAC	420
QY	3728	TCATCATTTGTCCCTATCTTGGGGAGAGCAGAGACACTACGTACCTCGGCCAGGTGG	3787
Db	421	TAATCATGTGTCCTATTTTGGAGAAATCCCAAMAAATATATGATGTATTTAGACAAGTAG	480
QY	3788	TCACCGGTGAGCGGTGCACATCTTCCGTGATTTATACATTCCTGTCCAAATTCGAAAGC	3847
Db	481	TACAGGGAAGACGATCACTCTCCGTAAATCTATACACTCTTCTTAAATTTGAAATG	540
QY	3848	AATACGAAGTCTCTATCTGAACCTCTGGACGCGTCTATAGTCTCTCAAGAAACAGC	3907
Db	541	AATATGAGCTTTTATTTATTTAAACCTTTACAGAGTGTATATGAAGTTTAAAAAACAAT	600
QY	3908	TGGAGAAATACGTATACCTTCACCTCAATGTCAAGTGAAGCAATCTGAAACACCGCTTTA	3967
Db	601	TAGAAATTAACCTTATATGACATTTATATGTATTTAAGGTATTTTAAATTCACATTTA	660
QY	3968	ATAACGAGAAATTTTCAAGACGCTCTGGAGAGCAGCTGATTCCTTTAAAGACCTGA	4027
Db	661	ATAAACGTGAATTTTCAAAAATGTTTAAATCAACATTTAATTCATTTAAAGATTTAA	720
QY	4028	CCCTCTCTAATAGTGTGAAGAACCCATACAGATCTCTCAATTAAGAAGAGAGGATA	4087
Db	721	CATCAAGTATTAATGTGTCTCAAGATCATATTAATTTCTTAATTAAGAAAAAAGATTA	780
QY	4088	AATTTCTGTCTAGTTTACACATATATCAAGAGCTCCATGCACACCCGATATCAATTTTCGTA	4147
Db	781	AATCTTAAAGACGTTAATAATTAATTAATTAAGATTCATTAAGATTCAGGATTAATTTTGCAA	840
QY	4148	ATGATGTCTCTGGGTATTTACAAAGTCTCTGAGCGAAATATACAAAGTCTGACTGCTA	4207
Db	841	ATGATGTCTCTGGATTTATTAATAATTAATTAATTCGAAAAATATTAATCAAGTTTATGATTTAA	900
QY	4208	TTAAAAAGTATATACAGGATTAAGCAAGGCGAGATGAATAAATATCGCCCTCTCGAATA	4267
Db	901	TTAAAAAATATATACAGACAAACAAAGGTAAATGAAATATACCTTCCCTTTTAAACA	960
QY	4268	ACATCGAAACCTCTTACACAGACGTGACAGACAATAATCGACTCTTTCGTAATTCACCTGG	4327
Db	961	ATATTTGACACCTTTATATTAACACAGTTATATATAAAAATTAATTTATTTGTAATTCATTTTGG	1020
QY	4328	AGGCCAAGCTCTCACTACATTTACAGCAAGAGCAGATGGGAAGTTAAATCAAGAGAGC	4387
Db	1021	AAGCAAAAGTCTTAATTAATATACATATGAGAAATCAAACTAGAGTTTAAATTAATAAGAAC	1080
QY	4388	TGAACCTACCTCAAAAACAATCCAAAGACACAGCTGGCAGATTTCAAGAAAAATAACAATTTGG	4447
Db	1081	TTAATTAATTAATAAACAATTCACAAATTTGGCAGATTTTAAAAAATAATACAAATTTTGG	1140
QY	4448	TCGGAATTTGCAGACCTGTCTACGAGTATATACACAACAATCTCCTGACCAAGTTTCTGT	4507
Db	1141	TTTGAATTTGATTTTATCAACAGATTTATACCATTAATTAATTTAGACAAAGTTCCTTA	1200
QY	4508	CCACCTGGCATGTGTTCGAAGAAACCTCGCCAAAACAGTGTCTGAGACATCTGTCGACGGCA	4567
Db	1201	GTACAGGATATGTTTTGAAGAAATCTTGCTTAACAACTGTTTATCTTAATTTACTTGATGAGAA	1266
QY	4568	ACCTTCAGAGGATCTGTAACAATCTCCAGCACCAATGTGTGAAGAAACAGTCCGCCACGA	4627

[illegible]

/codon_start=3
 /product="major merozoite surface protein precursor"
 /protein_id="BA02624.1"
 /db_xref="GI:391817"
 /translation="PLKTLSEESTIQEDNYASLENEKVLKLEKGLKONLNEKKKLS
 YLSGHLHLIAELKEVIRKNNYGNSSVNTDNNALLESYKFLPBGTVATVSS
 GSDLEQSPKRPASTHVGAESNITTSQWDEVDVLIIVJEGSEEDYDLDQYV
 TGEATPSVIDNLISLKEENEVLYLKPLGVRSLKOLENNMPEVNVKIDLSR
 FNRENKLVLESDLIPYKDLTSSNKKYKPKFENKERDKPLSSVNYIKDSIDPD
 NPNANDVLYKILSEKSKDPLSKIKKINROGENEYLFPLNIEFLYKTVNDKIDL
 FVHLEKLVNITYEKSNSVEVTKELNLYLTIDKLDKDEFKNNFVIAULSDYNIN
 NLTKFLSTGMVENLAKTVLSNLDNLGMLNI SOHOCVKKCPENSGCFRLDR
 EECKCLNTKQEDKVENPNPCNNENNGCDMDAKCTEEDSGSKKIKICECTKPDPS
 YPLFDGIFCSSNFGLISFLILMLILYSEFI"

BASE COUNT 694 a 198 c 240 g 504 t

ORIGIN

Query Match 18.48; Score 907.2; DB 3; Length 1636;
 Best Local Similarity 72.28; Pred. No. 1,7e-188;
 Matches 1179; Conservative 0; Mismatches 453; Indels 0; Gaps 0;

QY 3308 CTCTCTGAAGACCTCTCCGAGAGACATCCAGCCGAGATTAACCTCCAGCCGCTCG 3367
 DB 1 CTCTCTGAAGACCTCTCCGAGAGACATCCAGCCGAGATTAACCTCCAGCCGCTCG 60
 QY 3368 AGAAGCTCAAGCTCTCTGAGCTCGAGGCAAGCTGAGGACACCTGAACTGGAGA 3427
 DB 61 AAAAGCTTAAAGTATTAAGTAAATTAAGAGAAAATTAAGGATTAATTAATTAAGAAA 120
 QY 3428 AGAAGAGCTGAGCTCTCTGAGGAGCTGATCACTGAGCTGAGGAGCTGAGAGAG 3487
 DB 121 AGAAGAGCTGAGCTCTCTGAGGAGCTGATCACTGAGCTGAGGAGCTGAGAGAG 180
 QY 3488 TCATTAAGACAACTACACCGGCAATAGCCCAAGGAGATTAATCAAGCTGAATA 3547
 DB 181 TAAATTAAGACAACTACACCGGCAATAGCCCAAGGAGATTAATCAAGCTGAATA 240
 QY 3548 AGCACTGAGATCTTACAGAGAGTTCTGCTGAGAGAGAGATGAGCCGCTGGGTGT 3607
 DB 241 ATGATTAAGATCTTACAGAGAGTTCTGCTGAGAGAGAGATGAGCCGCTGGGTGT 300
 QY 3608 CTGAATCTGGCTCGAGACACTGAGAGAGTCTCAACCTAAGAGCTGATCTACTCATG 3667
 DB 301 GTGAAGTGGATCGGAGACACTGAGAGAGTCTCAACCTAAGAGCTGATCTACTCATG 360
 QY 3668 TCGAGAGCGAGTCCATATCAATTAACCATCTGAGAGAGTGGAGTGGATGACG 3727
 DB 361 TAGAGAGAGAGTCCATATCAATTAACCATCTGAGAGAGTGGAGTGGATGACG 420
 QY 3728 TCATCATTTGAGCTATCTGAGGAGAGAGAGAGTACATGAGCTGAGGAGGAGTGG 3787
 DB 421 TAATCATATGATCTATCTGAGGAGAGAGAGAGTACATGAGCTGAGGAGGAGTGG 480
 QY 3788 TCACCGGTGAGAGTGTCACTCTCTCGTATGATTAACATTTCTGCCAAAATCGAGAAG 3847
 DB 481 TAACAGAGAGAGTGTCACTCTCTCGTATGATTAACATTTCTGCCAAAATCGAGAAG 540
 QY 3848 AATACGAGAGTGTCTATCTGAGAGAGTGTGAGAGAGTGTCTGAGAGAGAGTGT 3907
 DB 541 AATACGAGAGTGTCTATCTGAGAGAGTGTGAGAGAGTGTCTGAGAGAGAGTGT 600
 QY 3908 TGAAGATTAAGCTGATGACCTTCAATGTCAGAGGAGATCTGAGAGGCGCTTTA 3967
 DB 601 TAGAAGATTAAGCTGATGACCTTCAATGTCAGAGGAGATCTGAGAGGCGCTTTA 660
 QY 3968 ATAAGAGAGAAATTTCAAGAGCTGTTGAGAGAGCTGATTCCTATTAAGAGCTGA 4027
 DB 661 ATAAGAGAGAAATTTCAAGAGCTGTTGAGAGAGCTGATTCCTATTAAGAGCTGA 720
 QY 4028 CTTCTCTAAGTGTGTGAGAGAGAGAGAGAGTCTCTCAATTAAGAGAGAGAGATA 4087
 DB 721 CATCAAGATTAATGTTGTGAGAGAGAGAGAGTCTCTCAATTAAGAGAGAGAGATA 780

QY 4088 AATTTCTGCTAGTTACACTATATCAAGAGAGTCCATGACGAGCATTAATTTCCGTA 4147
 DB 781 AATTTCTGCTAGTTACACTATATCAAGAGAGTCCATGACGAGCATTAATTTCCGTA 840
 QY 4148 ATGATGCTGGGTATTAACAAGATCTGAGAGAGAGAGAGTGAAGTGAAGTGAAGTGA 4207
 DB 841 ATGATGCTGGGTATTAACAAGATCTGAGAGAGAGAGAGTGAAGTGAAGTGAAGTGA 900
 QY 4208 TTAAGAGATATATCAAGATTAAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 4267
 DB 901 TTAAGAGATATATCAAGATTAAGAGAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGA 960
 QY 4268 ACATGAGAGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4327
 DB 961 ATATGAGAGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 4328 AGGCGAGAGCTCTCACTATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4387
 DB 1021 AAGCGAGAGCTCTCACTATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 QY 4388 TGAAGTACCTGAG 4447
 DB 1081 TGAAGTACCTGAG 1140
 QY 4448 TCGAGATTTGAG 4507
 DB 1141 TTGAGATTTGAG 1200
 QY 4508 CCAGTGGAGAGTGTGAG 4567
 DB 1201 GTACAGGATGAGTGTGAG 1260
 QY 4568 ACCTGAGAGAGAGTGTGAG 4627
 DB 1261 ACCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 4628 ATAGGAGAGTGTGAG 4687
 DB 1321 ATAGGAGAGTGTGAG 1380
 QY 4688 AACAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4747
 DB 1381 AACAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 4748 GGTGTGAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4807
 DB 1441 GGTGTGAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 4808 CATGAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4867
 DB 1501 CATGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 4868 CTAATTTCTGAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4927
 DB 1561 CTAATTTCTGAGAGAGAGATTAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 4928 AATGATGAGAG 4939
 DB 1621 AATGATGAGAG 1632

RESULT 25
 PRAMP8222 1636 bp DNA linear INV 29-MAY-2002
 LOCUS P. falciparum DNA for the precursor to the major merozoite surface
 DEFINITION proteins, C-terminal.
 ACCESSION D13359.1 GI:391796
 VERSION D13359.1
 KEYWORDS BGF-like domains; major merozoite surface protein precursor.
 SOURCE Plasmodium falciparum (isolate 82/2 and 82/7) merozoite, DNA.
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 1636)

QY	4668	AACAGGAAGAGATTAAGTCGCGTGGAGAACCCAAACCTTACTCTGCATGATAAACAATGGCG	4747
Db	1361	AACCAAGAGATTAAGTAAATGTGTTGAAATATCCAAATCTACTTGTAAACGAATAATATGGTG	1440
QY	4748	GGTGAGCGCCGACTCTAAATGCACCGAGGAAGACAGCGGCTCTAACGGAAGAAATATCA	4807
Db	1441	GATGTGATGACAGATCCCAAAATGTACCGAAGAAATTCAGATGACCAACGGAAGAAATATCA	1500
QY	4808	CATGCGAGTGTACTAAGCCCGGACTCCCTATCCACTCTTGACGGAGATTTTTCCTCAGCT	4867
Db	1501	CATGGAATGTACTAAGCTGATTTCTTATTCACACTTTTGATGATGATTTTTCGACATTCCT	1560
QY	4868	CTAATTCCTGGGCAATCTCTCTCTGTCGATGATCCATGCTGATCTGACACTTCATCT	4927
Db	1561	CTAATTCCTGGAATATCATCTTATTAATACATCATGTATATATATACAGTTTCATTT	1620
QY	4928	AATAGATGATG 4939	
Db	1621	AAAAATGTTAG 1632	
RESULT 26			
PMMSAMSA1			
LOCUS		3395 bp DNA linear INV 29.-JUL-1994	
DEFINITION		P. falciparum gene for major surface antigen (MSA-1).	
ACCESSION		X61930 S37357	
VERSION		X61930.1 GI:9917	
KEYWORDS		major surface antigen; MSA1 gene.	
SOURCE		Plasmodium falciparum	
ORGANISM		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
REFERENCE			
AUTHORS		Submitted (05-SEP-1991) U. Certa, F. Hoffmann-La Roche, Dept. PRTB,	
TITLE		Grenzacherstr. 124, CH-4002 Basel, SWITZERLAND	
JOURNAL		2 (bases 1 to 3395)	
REFERENCE		Olaifsson, P., Matile, H. and Certa, U.	
AUTHORS		Plasmodium falciparum: the repetitive MSA-1 surface protein of the	
TITLE		Ro-71 isolate is recognized by mouse antibody against the	
JOURNAL		nonrepetitive repeat block of Ro-33	
MEDLINE		Exp. Parasitol. 74 (4), 381-389 (1992)	
PUBMED		92275047	
FEATURES		1592091	
source			
		Location/Qualifiers	
		1..3395	
		/organism="Plasmodium falciparum"	
		/strain="Ro-71"	
		/db_xref="taxon:5833"	
		136..>3395	
		/codon_start=1	
		/evidence=experimental	
		/product="major surface antigen"	
		/protein_id="CA43932.1"	
		/db_xref="GI:9918"	
		/db_xref="SPTREMBL:Q25961"	
		/translation="MKTIIFLCSLSTFLPINTQCSTHSEYOELVKKLPALEDAVLTGSG	
		LHFHEKMLINDEGTTTGKASAGOSTGSCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG	
		SPSSRSNTLPKSNSSCAIIPADSDSDAKSYADLKRVNRYLFTTELKYPELFDTLT	
		NHMLTDLCDNIHGEYTLIDGVEINELLYKLNFFDLRAKINDCANDYQIIPNLKI	
		RAVELDLKRLVGVRRPLDNIKDNVSKMEDYIKNNQTTIANINELLEGSKITIDQKK	
		NADDEGKKRLYOAYDLSTYKNOLEAHNLIIVLEKRIPTLKNNKIKLEIDIKI	
		KTDKEPTTSSKPTMLDKNKKTIEHEKTIKETAKTIKPNIDSLFPLDELEYLRK	
		NKIVDTPKSDQPTKVOIKVPPNPIYVPLPTDINSLADNKNSGDLMNPT	
		KEKINKEIITDKKERKIFINNIKQIDLEKNNIHTKQNKRLLEDYEKSKDDEEL	
		EKEPEFNANNFDKVDVDKLFSARYTYNVERQKINNFSSNNNSVWVKAKLASL	
		EDVSLRIGISEKDFNNHYTTLTGLEADIKLTLEIKSENNKILEKNKGGTHSANAEL	
		EVSIVLIQVQKLYLKIETIDKRIELFLKNAQKDSIHVNIYKPNKKEPEYLIYIV	
		KKEVDKKEFIPKVKDKKQAVLSSITQPIVAASFTEDGGSHSTLSOSGETEVT	
		EEYGHGTTVTITPLPQPSDPKPKVKNVNSIEKSDNSQALTKIYTLKKLDEFLK	
		SYICHKTIIVSNSSMDOKLEVYNLTPPEENELKSCDPLDLFETQNNIPAMSLYDS	
		MNNDQLHFEFLYQEKEMIVYLHKIKENHNLKLEEDKQITGTSSTSSPENTVNTAQ	
		SATHSNQNDQSNASSNTNONGAVVSSGPAVVESSHDPITVLISNDLKIIVSLNNG	

		repeat_region		NKTVLPNPLITSTEMEKFEVENILKNNDDYFNDIDIKQFVKSNSKVITGLTETQKALN DEIKRKLPTLQSLSPDLNKKYKLJLDLRFNKKKELGQDKMOIKRLTLLKEQLESKLNS"	
		301..513			
		/evidence-experimental			
BASE COUNT	1512 a	439 c	434 g	1010 t	
ORIGIN					
Query Match	15.9%	Score 784.6	DB 3	Length 3395	
Best Local Similarity	57.9%	Pred. No. 1.6e-161			
Matches 1553	Conservative	0	Mismatches 1074	Indels 54	Gaps 7
QY	10	ATGAAATCATTTTCTCCTCTGTTCAATTTCGTGTTTTTAATCATCAATACTCAGTCGTG	69		
DB	136	ATGAAATCATCAATTCCTTTTATGTTCATTTCTTTTATTAATAAATACAAATGTGTA	195		
QY	70	ACCCAGAAATCCTATCAGAGAGCGTGTAAAGAAACGGAACGTTTGAAGATGCGCTCTT	129		
DB	196	ACACATGAAAGTTATCAAGAACTGTCTCAAAAAACAGAAAGCTTTAGAAGATGCAAGTATG	255		
QY	130	ACCGATACAGCGCTTCCACAGAGAAATGCGTGAATGAA-----	174		
DB	256	ACAGGTATATGTTTATTTTCATATAGAAAAAATGATCTTAAATGAAAGAAATTAATCA	315		
QY	175	-----GGAGCAGTGGCACGGCCGTTACAAACAGACAGCCGGTTCTAAAGG	222		
DB	316	AAAGTGCAAGTGTCACAAAGTGTCACAAAGTGTCACAAAGTGTCACAAAGTGTCACAAAGTG	375		
QY	223	TCTGTGCGTACGGGTGCGTCCGGTGGGTCTGTGCGCTTGCGGGTTCCGTGCGCTCCGGC	282		
DB	376	ACAAGTGTGTACAGTGTGTACAAAGTGTGTACAAAGTGTGTACAAAGTGTGTACAAAGTGTGTAC	435		
QY	283	GGCAGCGTGGCATCAGTGGGTGCTCAGTGGCAAGCGCGGTCCGGAAACAGTGCAGAAACC	342		
DB	436	GGTACAAAGTCCATCTCTCTGTTCAAAACCTTTACTCGTTTAAATACTTCAATCTGTGTGCA	495		
QY	343	AATCATCTGCACACTAGCGATTCGAGCGCCAGCGCAAGTCTCAAGCGCAGCCTCAAGACCGA	402		
DB	496	ATCCCTCCAGCTGATGCAAGCGATTCAGATGCTAATCTTACGCGATTTAAACACAGA	555		
QY	403	GTGAGAAACTATCTCCTCACTATCAAGAGGTGAAGTACCACAGTTGTTGCACCTCACT	462		
DB	556	GTACGAAATTAATCTTGTCTACTATTAATAAAGCTCAAAATVCCGAACCTCTTTGATTTAAC	615		
QY	463	AATCATATGCTGACACTGTGTGATTAACATTCATGCTTCAAAATATCTGATGAGCGTTAC	522		
DB	616	AATCATATGTTAATCTTGTGTGATTAATATTCATGTTTCAAAATATTAATGATGAGATAT	675		
QY	523	GAAAGATCATGAACCTCCTGTACAGTTGAATTTCTACTTCGACTTGTAAGGCCAAA	582		
DB	676	GAAAGAAATTAATGAATTTATTAATAAATTAACCTTTTATTTGATTTATTAAGAGCAAAA	735		
QY	583	CTGATGACGTTGGGCCAATGATGATTTGCAAAATTCATTCATTTGAAGATCAGAGCC	642		
DB	736	TTAAATGATGATATGTGCTAATGATTAATGTTCAAAATCCTTCAATCTTAAATTCGTGCA	795		
QY	643	AACGATGTGACAGTATTAAGAAGTTGCTTCCGATATGCAAGCGCTCGACAAATC	702		
DB	796	AATGATTAAGACGTACTTAAAAAACTTGTGTTGGATATAGAAAAACCTTAGACAAATAT	855		
QY	703	AAGCAATATGTGGAAAGATGGAAGATTAATTAATAAAGAAATGAAGACCATGAGAAC	762		
DB	856	AAAGATATATGTAGAAAAAATGGAAGATTAATTAATAAATAAACAACCATAGCAAT	915		
QY	763	ATTAAACGAGCTGATCGAAGAAATCCAAAAAGCCATAGACAAAAATTAAGATGCAACAG	822		
DB	916	ATTAAATGATTAATTAATGAAGAAATGAAGAAACCAATTTGTCAAAAATGAAGAAATGAGAT	975		
QY	823	GAGGAGAAAGAAAGAACTGTGACAGGCCAGATAGACGCTGTGCATCTATTAACAAACAG	882		
DB	976	GAGGAGAGAAAAAAATTTATACCAAGCTCAATATGATCTTTCTTATTTTACAAATAAACAA	1035		
QY	883	CTTGAAGAAAGCCATTAACCTCATAGCGTACTGGAAGAAAGCGCATAGACACCTCAAGAG	942		

Db 1036 TTAGAGAAACACATATATTTAATAGCGTTTGTAGAAAAACGATTTGACCTTTAAAAA 1095
 QY 943 AATGAAATATCAAGAACTGCTGACAGAGATTAAATGAATTAAGA---ATCCCTCCGCA 999
 Db 1096 AATGAAACACTTAAAGAAATCTTGAAGATATATGATTAATTTAAACAGATGCCGAAAA 1155
 QY 1000 GCCAATCTGGGAAACCCCTTAACACGCTGTGACAGAAAGAAAGATAGAGAGAC 1059
 Db 1156 CCCACTCTGAAAGTAAACCAATACTCTCTGATAGAACAAAAAATCGAGGAAAC 1215
 QY 1060 GAGAAAGAGTCAAGAGATCGGCAACCATATAGTTCAAGTCAAGTCTCTCTTACT 1119
 Db 1216 GAGGAAAAATTAAGAAATGGCAAAATCTTAATTAATTAACATGATGATTTATTTACT 1275
 QY 1120 GATCCCTTGGAGTGAAGTCTACTTGTAGAGAGAGATAGAAATAGACATCTCCGC 1179
 Db 1276 GATCCACTTGAATTAAGATTTATTTTAAGAAAAAATTAAGAGTTGATGACACT 1335
 QY 1180 AAGTGA-----GACAAAGAAATCAACCGAACCTTAATGATTTCCCAATGTGTG 1230
 Db 1336 AATCAACAGATCTCAAGAAATCTGTTCAAAATACCAAAAGTCTTAATCCAAATGTATT 1395
 QY 1231 ACCTACCCCTGCTTATTAACGATATCAACA-----CGCTCAACGAGCTCAAT 1281
 Db 1396 GTATATCTTTACCACTGCTGATATTCATTAATTCATTAAGCTGACATTAATGATTAAT 1455
 QY 1282 AGCTGCGTGAATGATTAACCCCTTGATTAACGAAACCCCTTAAGAAATATCTAC 1341
 Db 1456 TCATATGTGATTTAATGAATCTGATTAAGAAAAAATTAATGAATAATTAATTA 1515
 QY 1342 ACAGACATAGAGAAAGAAATTTATCAACGAATCAAGAGAGATCAAAATTTGAGAG 1401
 Db 1516 GATTAATAGAGAAAGAAATTTATCAATTAATTAATTAATTAATTAATTAAGAA 1575
 QY 1402 AAGAAATTTAGAGTGAAGAAAGTGAAGACCGCAAGAAAGTCTTAACGATATC 1461
 Db 1576 AAAAACTTTATACACAAAGAAACAAATTAATTAATTTCTTGAAGTTTGAAGATCA 1535
 QY 1462 ACTAAGAGATGAAAAAGCTGCTGAACGAGATCTATGATTTCAAAATTCACAAATATCATC 1521
 Db 1636 AAAAAAGATATGAGAAATTAATTTGAATAATTTATGAATAATTAATTAATTTT 1695
 QY 1522 GACCTGACCACTTCGAAAGAAATGATGGGAAACGCTACTTTACAAAGTGGAGAACTG 1581
 Db 1696 GACAAAGATGCTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1755
 QY 1582 ACACACATATATACCTTTCGATGAGATCTATGAGATCTATGATCTGAGAGCTCAC 1641
 Db 1756 ATATATATATATATATTTTATCTCTTAATTAATTTCTGTATATATGTTCAAAATTTAAA 1815
 QY 1642 AAGCTCTTAAGTATATGAGAGATATTCCTGCGAAACATTTGTGTGAGAAAGACTA 1701
 Db 1816 AAGGCTCTTTCATATCTTGAAAGATTAATTTCTTTAAGAAAGAAATTTCTGAAAAAGATTT 1875
 QY 1702 AAGTATTAACAAGATCTCATTAAGTGAAGTGAAGACGATCGAGACGCTTTGTGACAC 1761
 Db 1876 AATCATATATATCTTTGAAAACTGGCTCGAAGCTGATTTAAAAAATTAACAGAAAG 1935
 QY 1762 ATTAAGAGAGTGAAGAACGTTGTTGAGAGAGATTTCAAAAGCGAAATTAACCA 1821
 Db 1936 ATTAAGAGTGAAGAAACAAATTTCTAGAAAAAATTTTAAGAGACTAACACATTCAGCA 1995
 QY 1822 GATGAGAAAGTCTCGAGTCTCGATATTTGTTAAAGTCCAAAGTGAAGAGTGTCTCTC 1881
 Db 1996 AATG---CTTCTTAGAGATCTGATATTTGTAATAATTAACAGTAAAGTTTATTA 2052
 QY 1882 ATGAACAAGTTTATGACTCAAGAGACTCAACTCTTGTGAAGACGTTGAGTTAAA 1941
 Db 2053 ATTAAGAAATTAAGAGACTTAAGAAAGATGAATTTATTTTAAAAAATGACAACTAAAA 2112
 QY 1942 CATATATATCATGTCGAGATATGTTAATAGCAGAGAGATAGAGAAACCATCTACTCTC 2001
 Db 2113 GATAGATTCATGTACCAATATTTATTAACCAAAATTAACCAAGAACCATATATTTA 2172

QY 2002 ATGTACTCAAGAAAGATAGACAAACTGAAAGTGTTCATGCCAAAGTCGAGACCTG 2061
 Db 2173 ATTGTATTAAGAAAGAGATGATTAATTAAGAAATTTATACCAAAAGTAAAGACATG 2232
 QY 2062 ATCAACGAGAGAGAAAGAAACATTAATACTGAAGACAGTACAGATACCTCCAGCTTC 2121
 Db 2233 TTAAGAA--AGAACAGCTGTCTTATCAAGTATTCACAACTTTAGTTGACGAAGG 2290
 QY 2122 ACAGAGAGAGATTAACCGGACAGGCTACACCAAGCCCGACAAAGCGCGGTAGCT 2181
 Db 2291 AATCAACTGAAAGTGGG-GGTCACTCCACACACATTTATCCAAATCAGAGAAACAGAA 2349
 QY 2182 CTCGAAGCGATGCGTGCAGAGCTCAAGCAAGAGACAGAGAGGACAGCTCCAGTG 2241
 Db 2350 GTAAAGAAAGAAACAGTAGACACACAAAGCGTAAACATTAATTAATTAATTAATTA 2409
 QY 2242 CCAAGTCCGCTTCAGAGCTAAAGCTCAAGTCCCTACACACAGCTCCTGTGATTAAC 2301
 Db 2410 CCATCACACCAAAAGAGTAAAGTTGTAATAATTAATTAATTAATTAATTAATTA 2469
 QY 2302 AAGACGAGATGTCAGCAACTGACACTGAGTCTGAGAGCTATGATGATTCGATTAAC 2361
 Db 2470 AATTCACAGCTTGCACAAACAGTTTATCTAAAGAAATTAATTAATTTTAATTA 2529
 QY 2362 TCCTACATCTGCACAAATATATCTGCTCTCACAGCACTATGAAACGAGATTTT 2421
 Db 2530 TCATATATATGATCATTAATTTATTTAGTCAAACTGTATGAGCAAAATTAATTA 2589
 QY 2422 AATCACTACAGATTAACCAAGAGAGAGAGATTAATGCTCCCTGTATGATCCAGGAC 2481
 Db 2590 GAGGTATATATATCTTCTCCAGAGAGAAAGAAATTAATTAATTAATTAATTAATTA 2649
 QY 2482 CTGCTGTCAATATCCAGAAACATTCCTGATATATCTATGTTGCTATGCTCAAC 2541
 Db 2650 TTATTTATTTATTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2709
 QY 2542 AATTCCTCTCACTGTTATGAGATATATGAGAGAGATGCTGCAACTGTAT 2601
 Db 2710 AATGATTTACAACTCTCTTTTGAATTAATTAATTAATTAATTAATTAATTA 2769
 QY 2602 AATCTCAAGACACAGCAAGATTAAGAACCTTGAGAGA 2642
 Db 2770 AATCTCAAGAGAGAAATCAATCAAAATTAATTAATTAATTAATTAATTA 2810

RESULT 27
 PPAP190A 3518 bp mRNA linear INV 26-APR-1993
 LOCUS P.falciParum merozoite 190 kd precursor protein (p190) mRNA, 5'
 DEFINITION
 end.
 ACCESSION M35727
 VERSION M35727.1 GI:160549
 KEYWORDS p190 gene; variable surface antigen.
 SOURCE Plasmodium falciparum (individual isolate RO-33 Ghana) cDNA to mRNA.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 3518)
 Certu, U., Rotmann, D., Matile, H. and Reber-Liske, R.
 A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats
 EMBO J. 6 (13), 4137-4142 (1987)
 JOURNAL MEDLINE
 MEDLINE 8816657
 PUBMED 3327688
 FEATURES
 source
 Location/Qualifiers
 1..3518
 /organism="Plasmodium falciparum"
 /isolate="RO-33 Ghana"
 /db_xref="taxon:5833"
 1..>3518
 /product="p190 protein"
 /note="RO-33 mRNA"

Query Match	Best Local Similarity	Matches 1511; Conservative	14.4%; Score 711.8; DB 3; Length 3518; 56.8%; Pred. No. 1.66-145; Indels 76; Gaps 8;
QY 10	ATGAAATCATTTTCTCTCTGTTCAATTCGTTTTTATCATCAATATCATGTCGCTG	69	
DB 337	ATGAAGATCATATCTTTTATGTTCAATTCGTTTTTATTAATATCAATATGTTGA	396	
QY 70	ACCCAGCAATCCTATCAGAGCGTGGTTAAGAAACGCTTGGAGATGACCGCTCTT	129	
DB 337	ACACATGAAAGTTATTCACAGAACTGTCTCAAAAACTAGAAAGCTTATGAAAGATGCACTATTG	456	
QY 130	ACCCGATACAGCGCTTCTCAGAAAGAGAGATGTGTGCAATGGAAGGACACAGTGGCAGC	189	
DB 457	ACAGGTTATAGTTTATTTCAAAAGSAAAAAATGTAATTAAGAGATGGAGCAAAATATCTAA	516	
QY 190	GCCGTTAACACACGACACCCGGTTCTTAAGAGGTCGTGGCTAGCGGTGGCTCCGGTGGG	249	
DB 517	GTTGTTGCAAAAGCCTGCAGATGCTTAAGTACTCAAACTGCTAAAAAATCTCCAGGTTCT	576	
QY 250	TCTGTGGCCTCTGGGGGTTCCTCGCTCCGCGGACGCTGGCATCAGGTGGCTCAGTG	309	
DB 577	ACAGTACTTTCAGTACTGCAAGTACTTAAGTGTCTTAATATCTCCACAGGTGCTGCA---	633	
QY 310	GCAAGCGCGGTTCCGGGAACAGTGAAGAAACAATTCATCTGACCACTAGCGATTCC	369	
DB 634	-----AATCCTTGAGATGATTCAAGTGAATCA 660		

OY	370	GAGCCAAAGTC	TA	GCCGCACTC	AA	GCACAC	CGAGTGA	GA	AACTAT	CTCTCA	TATCA	G	423
- Db	661	GATCGTAATCT	T	ACGCTGATTT	AAAA	CA	TAGAGTTC	CAAA	ATTTACT	GTGTC	ATTTAA	720	
OY	430	GAGCGAAGTAC	CCCA	CGGTTC	CGAC	CTC	ACATAT	CT	GAC	ACTGTG	GTATAC	489	
Db	721	GAACCAAAAT	T	CCCGA	CTCTT	TGATTTA	ACCAAT	CA	TATGTTA	ACTTGTG	TATAT	780	
OY	490	ATTATGGCT	CA	AAATAT	TC	ATGAT	TGAC	GGTTA	CGA	AGATCA	ATGAC	549	
Db	781	ATTCATGGT	TTCA	AAATTTA	T	AT	TGAT	GGAT	T	ATGA	GAATTA	840	
OY	550	TTGAATTT	CTA	CTT	CGACT	CT	CTA	AGG	CCAA	CTGA	TGAC	609	
Db	841	TTAACTTTA	TTTTG	ATTTA	T	TA	AGG	CAAA	TTTAA	TGA	TGTCT	900	
OY	610	TGTCAAAT	T	CCATTC	CA	TTTTG	AA	ATC	AG	CCCA	AGATG	669	
Db	901	TGTCAAAT	T	ACCTTT	CA	TTCTT	AA	AAATTT	CG	TCA	ATTTGA	960	
OY	670	GTCCTCGAT	AT	TCGCA	AGCCT	CT	CGAC	A	AT	TC	AG	729	
Db	961	GTGTTCCGAT	AT	TAGAAA	ACCA	TTAG	ACTTTAT	T	AA	GA	TATGTG	1020	
OY	730	TATAT	TA	AAAA	GA	AT	TAG	AA	GAC	CTC	AG	789	
Db	1021	TACAT	TTAAAA	AAAA	TTAA	AC	ACCA	T	AG	CAAA	TTTAA	1080	
OY	790	AAGAC	CTAT	AGCA	AAAT	T	AA	ATG	CA	AC	AG	849	
Db	1081	AAACAT	TGAT	CAAA	TTA	TGA	AT	TG	CA	AA	AG	1140	
OY	850	GCCCA	GTAG	CA	CTGT	CCAT	CT	AT	AA	CA	CA	909	
Db	1141	GCTCAAT	AT	TGAT	CTTTT	ATTT	ACA	ATTTAA	CA	ATTTCA	GAAG	1200	
OY	910	GTCAT	CGA	GA	AGCC	CA	TAG	ACCC	CT	CA	GA	969	
Db	1201	GTTT	TAGAAA	AG	CTAT	TG	AC	CTTTA	AAAA	AAAA	CTA	1260	
OY	970	AAGAT	TAT	GA	AAAT	T	TA	GA	AT	CT	CT	1027	
Db	1261	GATAT	AG	TAAAT	T	TA	AT	TAG	AT	TG	CG	1320	
OY	1028	TGCTG	AC	AG	AC	AG	AC	AG	AT	AG	AG	1087	
Db	1321	TCCCTG	AG	AT	TAG	AAAA	AG	GT	CG	AG	AG	1378	
OY	1088	CCAT	TAG	TTCA	CA	AT	GA	TTCT	CT	CT	CT	1147	
Db	1379	CTAT	TAA	AAATTTA	AC	AT	TG	AT	AG	TTT	AT	1438	
OY	1148	GAGAGA	AG	AT	TAG	AAAT	T	AG	AT	CT	CG	1198	
Db	1439	GAGAAAAA	AT	AAAAA	AG	TTG	AT	GA	T	GA	CA	1498	
OY	1199	CAAC	GA	AACT	AT	GA	TAT	CC	AA	TG	TG	1258	
Db	1499	AAAA	TAC	CA	AAAG	TTCT	CT	TAT	CA	AAAG	TG	1558	
OY	1259	ACAA	-----	CG	CT	CA	AG	AG	CT	CA	AT	1309	
Db	1559	ATTA	AT	CTA	T	AG	CT	CA	GA	T	AT	1618	
OY	1310	ATTAT	TAG	AA	GA	AA	GA	AA	CT	CT	TAT	1369	
Db	1619	CTAA	GA	AAAAA	AT	T	AA	T	GA	AA	AA	1678	
OY	1370	ACGA	AT	CT	CA	AG	AG	AT	CA	AA	AT	1429	
Db	1679	ATA	CA	AT	T	AAAAA	CA	AA	AT	T	GA	1738	

OY	1430	ACGAGACCCGACGAGCAAAAGCTTAACAGATATCTCTAAAGAGTATGAAAAGCGTGAAGC	1489
Db	1739	ATAAAAATTAATCTGGAAGGATTAATGAAAAAGTCAAAAAGATTAATGAGACATTACTTGAA	17989
OY	1490	AGATCTATGATTCCTCAAAATTCACAAATPACATCGACCTGACCAACTTCGAGAAATGATGG	1549
Db	1799	AATTTTATGAATGAATAATTAATTAATTAATTTTAAACAAGATGTCGTAGATAAATATTTCA	1858
OY	1550	GAAACGGTACTCTTACAAAGTGAGACAACCTGACACACATATATACCTTTGCATCTTATG	1609
Db	1859	GTGCAGATATATACATATTAATTAATGTAAGAAACCAAGATATATATTAATTTTCATCTCTA	1918
OY	1610	AGAAATCAACCATATATCTTGAGAAAGCTCCAAACCTCTTAAGTATATGAGGAGCTATT	1668
Db	1919	ATAATTCGTATATAAAGTTCCAAAATTTAAAAAGGCTCTTTCATATCTTGAAGATATT	1978
OY	1670	CTCTCGCGAATCTGTTGTGGAGAAAGAACATAAGTATTACAAAGATCTCATACATATAGA	1729
Db	1979	CTTTAAGAAAAAGAAATTTTCGAAAAAGATTTTAATCATTTATTTATCTTTGAAAACTGGC	2038
OY	1730	TCGAAAACGAGATCGAGACGCTTGTGGAAACATTAAAGAGATGAAGACAGTTGTTG	1789
Db	2039	TCGAAAGCTGATATATAAAATTTAAACAGAAATTAAGAGTATGTAAGAAACAAATTTCTAG	2088
OY	1790	AGAAAGATATCAAAAGACGAAATTAACCCATATGAGAAAGATCCTCGAGAGCTCCGATA	1849
Db	2099	AAAAAATTTTAAAGAGCTACACATTTACGCAAAATCTTTC - - -CTTGAAGATATCTGATA	2155
OY	1850	TTGTAAAGTCCAAAGTCACAAAGGTGCTCCTCATGAACAAGATTGATGACTCAAGAGA	1909
Db	2156	TTGTAAATTTACAGTACGCAAAAGTTTATTAATTAATTAATTAATTAAGACGCTTAACAAAGA	2215
OY	1910	CTCAACTCATTCGTGAGAGACGTGGAGTTAAACATATATACATGTGCCAATAGTTATA	1969
Db	2216	TAGATATTATTTTAAAAAATGCAACAATAAAGATGTATTCATGACCAAAATATTATTATA	2275
OY	1970	AGCAGAGAAATTAACGAGAACCATATCTACTCTACTACTCAAGAAAGATATAGCAAC	2029
Db	2276	AACCAAAATTAACCGAAACCATATTATTATTATTATTAATTAAGAAAGATGATTAAT	2335
OY	2030	TGAAAGTGTTCATCCCAAGGTGAGAGCGCTGATCAACGACAGAAAGAAATTAATAA	2089
Db	2336	TAAAGATTTATTAACCAAAAGTAAAGACATGTTAAAGAAAGAACAGCTGCTTATACAA	2395
OY	2090	CTGAAGCAGCTCAGATTAAGTCCGAGCGTTCCACAGAGAGAGATTAACCGGACAGGCTA	2149
Db	2396	GTATTACACAACTTTTAGTTG - - - - -CAGCAAGCAAAACAACTGAAGATGGGG	2443
OY	2150	CCACCAAGCCCGGACACACAGCGCGGTTCAAGCTCTCAAGCGCATACCGTCAAGCTCAAG	2209
Db	2444	GTCACTCCACACACATTTATCCCAATCAAGAGAAACAAAGATTAACAGAAACAGAA	2503
OY	2210	CACAAGACACAAAGCAGAGCACAGCTTCACAGTGCAGCTGCCCTCCAGAGGCTTAAGCTC	2269
Db	2504	AAACAGTATGGACACACAAACAGGTAACTAATCATTTACCAACCAAAAGATTAAGATTG	2563
OY	2270	AAATGCTACACCAACGAGCTCTGTGATATACAAGACGAGATGTCAAGCAATGAGCT	2329
Db	2564	TTGAAATTCATATAGACATTAAGAGTATATGACATTTACAAAGCGCTTGACAAAAACAGTTT	2623
OY	2330	ACCTTGAAGAGCTTATAGAGTCTCTGATATCATCTTACATCTGCCACAATATATCTCG	2389
Db	2634	ATCTAAAGAAATTTAGATTTTAACTAAATCAATATATATGTCATATATATTTTGG	2683
OY	2390	TCTCTCACAGCATATGAAGAGAGAGTCTTAAACAGTACAAAGATTAACCAAGAAAGG	2449
Db	2684	TATCAAACTCTAGTATGACCAAAATTTATTAAGGTATATATCTTAC - - -TCCAGAG	2740
OY	2450	AGAGTAATCTGCTCTTGTATCCACTGAGCGCTGTTCTCAATATCCAGAAACATTC	2509
Db	2741	AAATTAATTAATTAATCATGTGATGATGATATTAATTTATTAATTAATTAATTAATTAACATAC	2800
OY	2510	CCGTTATGTATCTATGTGATAGCCTCAACAATTTCTCTCAACTGTTCATGAGAGA	2569

DB	2801	CGCGATGATGATTCATTATATATGATTAATGATTAAGCAACATGATTTACACATCTCTTTTGAAT	2860
QY	2570	TATATGAGAAGAGATGGCTGCAACCTGTATAACTCAAGACAGCAAGATTAAAGA	2629
DB	2861	TATATCAAAAGAAATGATTTATTTATTTACATAACTAAAGAGAGAAATACATCAAAA	2920
QY	2630	ACCTCTTGAGAGAACTAA	2648
DB	2921	AATTATTAGAGAGCCAAA	2939
LOCUS	PP190G	3518 bp	DNA
DEFINITION	Plasmodium falciparum surface antigen P190 gene fragment.	linear	INV 29-NOV-1987
ACCESSION	Y00087		
VERSION	Y00087.1	GI:9924	
KEYWORDS	P190 gene; surface antigen.		
SOURCE	Plasmodium falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Certa, U.		
TITLE	1 (bases 1 to 3518)		
JOURNAL	Direct Submission		
REFERENCE	Submitted (03-NOV-1987) Ulrich C., Hoffmann-La-Roche, Department		
AUTHORS	2FE, Grenzacher Str. 124, 4002 Basel, Switzerland		
TITLE	2 (bases 1 to 3518)		
REFERENCE	Certa, U., Roemman, D., Matile, H. and Reber-Liske, R.		
AUTHORS	A naturally occurring gene encoding the major surface antigen		
TITLE	precursor P190 of plasmodium falciparum lacks tripeptide repeats		
JOURNAL	EMBO J. 6 (1987) in press		
FEATURES	Location/Qualifiers		
SOURCE	1..3518		
	/organism="Plasmodium falciparum"		
	/strain="Ro-33"		
	/db_xref="taxon:5833"		
	/clone="Ro-33/1"		
	337..>3518		
	/note="P190 surface precursor (AA 1-107) (3518 is 2nd base in codon)"		
	/codon_start=1		
	/protein_id="CA68280.1"		
	/db_xref="GI:9925"		
	/db_xref="SWISS-PROT:P19598"		
	/translation="MKIIFFLCSELPFPIINTQCVTHESYOELVKLEALDEAVLTVGYS		
	LFQEKRMALKGANTOVYAPADAVSTQSAKPNPGATVPSTGASTKGISPGANPSS		
	DDSSDAKSYADLKHRYONLFTPEKLYKELPDNLNHHMLTDNLHGFFLYLDGE		
	EINLILYKINFPDLRLKALDNCVANCYCPMLKIRANDLGLKLVGYSRKLDE		
	IKDVGAKMEDITKRNKTTIANINELIESGSKTTIDQKNANDEBKRLIQAQYDLPITY		
	NKQAGAHNLISLEKRTIDTLKKNENIKLEEDIDKIRIDAEPTGVNOLISLRLEK		
	ESRBEKEIKETAKTIKENIDILFTDPLELEYLRKKKQVDPKSDQPKSVQIPVW		
	PYNGVIVPLDIDIHNSLADNDKNSGIDLMNPTKEIKINELITDNKEKKIFINMI		
	KKOIDLEKKNINHRKQNKLLIEDYEKSKQYEELDEPKEMNNENKQVVDIKFS		
	ARTAYVNEKQRYNNKFESSNSNSVYNNQKIKALSYLDYSIRGKISRPKHYYTLTKT		
	GLEADIKLLEIKSSEKKILLENKFKGLHSANLSLEYSDIVKIQVQKVLLIKIEDL		
	KRIELFKMAQKSDSHVPIVNPVQPNKPEPYLLVLKEDVKELEFIPKXMDKKEQD		
	AVLSISITQPLVAASEETLEDGSHSTHTLDSQGETEYETETETVHTTVAITTPRKE		
	KVVENSLIEKNSDNOALITVYLKIDELFKYSICHYLVIVENSMDOKLELVNML		
	TPRENEILKSGDRLLDLEFNIONNIPAMYSIVDSMNNDLOHEFELIYKEMIIYLKLEK		
	ENHITKLELEPKQTTGTSSTSSPCGNTTVYNTAQAQSTHSNSQDQSNASTNQNQVAVS		
	SGPVAVESHDPLTVLSISNDLKGIVSLNLGNTKVPNPLTISTEMERVENILKTI		
	MIPIFNDDIKQFVNSKSVITGLNETQKNALNDEIKRLKDTLQISFDLYNNYKRLKDR		
	LENKKKEIGDQKMOIKIKTLLEQLEKINS"		
BASE COUNT	1570 a 430 c 437 g 1081 t		
ORIGIN			
Query Match	14.4%	Score 711.8;	DB 3; Length 3518;
Best Local Similarity	56.8%;	Pred. No. 1.6e-145;	
Matches 1511: Conservative	0;	Mismatches 1072;	Indels 76; Gaps 8;
10 ATGAAATCATTTTCTCTCTGTTTCATTTTCTGTTTATCATCAATACAGCGGTG	69		

Db 337 ATGAAGATCATATCTTTTATGTTTCATTTCTTTTTTTTATTATAAATACAAATGTGA 396
 QY 70 ACCGAGATCCTATCAGAGCTGTTAGAACTGGAAGCTTTGGAACATGCCCTT 129
 Db 397 ACACATGAAAATTATCAAACTGTGCAAAAACATAGAAAGCTTTAAGATGAGATTGG 456
 QY 130 ACCGATACAGCCCTGTCAGAGAGAGATGCTGAATGAAAGGAGAGAGATGCGACG 189
 Db 457 ACAGGTTATAGTTTATTTCAAAAGGAAAAATGTTATTAAGATGAGACAAATACTCA 516
 QY 190 GCCGTTACACACACACACCCGGTCTAAAGGCTGTGCGTACGGGTGCGGCGG 249
 Db 517 GTTGTGCAAGCCCTGCAATGCTGTAACTCAAAAGTACTCAAAATCTCCAGGTGCT 576
 QY 250 TGTGAGCCCTGTCGGGGTTCGGCTCCGGGCGAGGCTGGCATCAGGTGGCTAGAG 309
 Db 577 ACAGTACCTTCAGGTAAGTCTCAAGTACTAAAGGCTGTAAAGATCTCCAGGTGCTGA --- 633
 QY 310 GCAGCGGCGGTTCCGGGAACAGTCGAAAGACCAATCATCTGACAACTGACGATTC 369
 Db 634 -----ATCCTTCAGATGATTCAGGTATTC 660
 QY 370 GACGCCAGTCTCAGCCGACCTCAAGCACCGAGTGAAGTATCTCCTCACTATCAAG 429
 Db 661 GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTAATGATTTATTTAA 720
 QY 430 GAGCTGAAGTACCCAGTGTTCGACCTCAATCATGCTGACAGTGTGATAC 489
 Db 721 GAATCAAAATATCCGAACTCTTGTATTAACCAATCATGTTAATCTTGTGTATAT 780
 QY 490 ATTCAATGCTCAAAATATCTGATGATGACGTTAGAGAGATCAAGTACCTGTGCAAG 549
 Db 781 ATTCAATGCTCAAAATATTTAATGATGATGATGAAGAAATTAATGATTTATTTAA 840
 QY 550 TTGAATTTCTACTGACTCTGAAGGCCAACTGAATGACGTTTGGCCATGACTAT 609
 Db 841 TTAACCTTTTATTTGATTTATTAAGAGCAAAATTAATGATGATGATGATGATTTAT 900
 QY 610 TGTCAAAATTCATCAATTTGAAGATCAGAGCAACGAGTGGAGCTATTTGAAGATTTG 669
 Db 901 TGTCAAAATTCCTTCATGATTTAAATTCGTGCAAAATGATTTGACTACTTAAAAACTT 960
 QY 670 GTCTTCGATATCGAAGCCTCTCGACAACATCAAGACATGTGGGAAAGATGAAAT 729
 Db 961 GTGTTCGATATGAGAAACCATTAAGCTTTATTAAGATTAATGAGAAATATGAAAT 1020
 QY 730 TATATTTAAAAAATTAAGAACATCGAAGCATTTAACGAGCTGTGCAAGATTCAAA 789
 Db 1021 TACATTTAAAAAATTAACCAATGCAATGCAAAATTAATGAAATTAATGAAGAAAGTAA 1080
 QY 790 AAGACCATGACAAAAATTAAGATGCAACCAAGGAGGAAAGAAAGATTTGTCCAG 849
 Db 1081 AAAAAATTTGATCAAAATTAAGATGCAATTAATGAAGAAAGAAAAAATTTATRCAA 1140
 QY 850 GCCGATACGACCTGCTCATATTAACAACAGCTTTGAAGAGCCCATACCTCACTCAG 909
 Db 1141 GCTCAATATGATCTTTTATTTACATAAACAATTTACAGAAAGCACATATTTAATTAAC 1200
 QY 910 GTACTGAGAGAGCGATAGACACCTCAAGAGATGAATAATATCAAGAACTGCTCGAC 969
 Db 1201 GTTTTGAAGAAACGTATGACCTTTAAAAAATTAAGAAACATTAAGAAATTTAGAA 1260
 QY 970 AAGATTAATGAATTAAGATCTCT---CCGCGAGCCCAACCTCTGGAGACACCCTACACGC 1027
 Db 1261 GATTAAGATTAATTAATTAAGATGCGAAAAACCCACTACTGAGATGAACCAATCTCTC 1320
 QY 1028 TGCCTGAGAGAGAGAGATGAGAGAGAGAGAGAGATGCAAGAGATGCGCAAAA 1087
 Db 1321 TCCCTGAGATTTAGAAAAAGTCGAG - ACAGAGAAAAAATTAAGAAATTTGCCAAA 1378
 QY 1088 CCAATTAAGTTCAACATGATCTCTCTTTTACTGATCCCTGAGCTGAGTACTACTGA 1147
 Db 1379 CTATTAAATTTAAACATGATGATGATTTATTTACTGATCCACTGAATTAATTAATTA 1438

QY 1148 GAGAGAGATTAAGATTAATGACATCTCCGCCAAAGTCGA -----GACAAAGAAAT 1198
 Db 1439 GAGAAAAAATTAAGATTTGATGTAACCTTAATACACAGATGCTTACGAAATCTGTTT 1498
 QY 1199 CAACCGAACCTTAATGAATATCCCAATGCTGTGACGTACCTCTGCTCTTAATACGATATGA 1258
 Db 1499 AATATCCAAAGTCTCTTCCAAATGATGATGATGATGATGATGATGATGATGATGATG 1558
 QY 1259 ACAA-----CGCTTCACAGAGCTCAATAGCTTGGTGTGACTGATTACCCCTTG 1309
 Db 1559 ATAAATTCATTAAGCTGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 1618
 QY 1310 ATTAATGAGAGAACCTCTAAGATATGCTACACAGACATGAGAGAGAAAGTATTAATCA 1369
 Db 1619 CTAAAGAAAAAATTAATGAAAAATTTATTAAGATTAATTAAGAAAGAAAAATTTATTA 1678
 QY 1370 ACAGAAATCAAGAGAGATCAAAATTTGAGAGAGAGAAATTTGAGAGTGAACAGAAAGTT 1429
 Db 1679 ATAACTTTAAAAACAAATTTGATTTAGAGAAAAAACAATTAATCAACAAAAAGAACTAA 1738
 QY 1430 ACAGAGACCGCAGCAAAAGCTTAACGATATCTTAAGAGTATGAAGAGTCTGCAAG 1489
 Db 1739 ATAAAAAATTTACTTGAAGATTAATGAAAAAGTCAAAAAGGATTAATGAAGAAATTTGAAA 1798
 QY 1490 AGATCTATGTTTCCAAATTCACAAATACATGACCTGACCAACTTCGAGAAAAATGATG 1549
 Db 1799 AATTTATGAAATGAAATTAATTAATTAATTTTAAACAAAGTGTGTAAGTAATTTTCA 1858
 QY 1550 GAAACGGTACTCTTCAAAAGTGAAGAACCTGACACACCATATATACCTTTGATCTATG 1609
 Db 1859 GTGCAATATATACATTAATTAATGTTGAAACAAAGATATATATTAATTAATTTTCTCTGA 1918
 QY 1610 AGATTTCTAAGCTATATCTTGAAGAGCTCAACCAACTCTTAAGTATATGAGAGACTATT 1669
 Db 1919 ATAAATCTGTATTAATTAATGTTCAAAATTAATTAATTAATTAATTAATTAATTAATTA 1978
 QY 1670 CTCTGCGGAACATTTGTTGAGAGAAAGTAAAGTATTAACAAGTATCTCATTAAGTAAGA 1729
 Db 1979 CTTTAAGAAAGAAATTTCTGAAAAAGATTTTATATATTAATTAATTAATTTGAAACCTGGCC 2038
 QY 1730 TCGAAAAAGAGATCGAGACCTGTTGAGAACATTAAGAGATGAAGAGATGTTG 1789
 Db 2039 TCGAACCTATATTAATAAATTAACAGAAAGATTAAGAGATGAGTAAACCAAAATTTCTG 2098
 QY 1790 AGAAGAGATTTCAAAAGAGCAAAATTAACCAAGATGAGAAATCTCGAGAGTCTCGATA 1849
 Db 2099 AAAAAATTTTAAAGGACTAACACATTCAGCAATGCTTC ---CTTAAGAGTATTCGATA 2155
 QY 1850 TTGTTAAAGTCCAAAGTGAAGAGTCTCTCATGAACAAGATTTGATGAACCAAGAGA 1909
 Db 2156 TTGTTAAATTTACAGATACAAAAAGTTTATTAATTAATTAATTAATTAATTAATTAATTA 2215
 QY 1910 CTCAACTCATTTGAGAGAGAGTGAAGTAAACATTAATTAATTAATTAATTAATTAATTA 1969
 Db 2216 TGAATTAATTTTAAAAAATGACACACTAAAGATGATGATGATGATGATGATGATGATGAT 2275
 QY 1970 AGCAGAGATTAAGAGAGACATTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 2029
 Db 2276 AACCAACAAATTAACCGAGAACCATTAATTAATTAATTAATTAATTAATTAATTAATTA 2335
 QY 2030 TGAAGTGTTCATGCCCCAAGTGAAGGCTGTCACAGAGAGAGAGAGAAATTAATTA 2089
 Db 2336 TAAAGATTTTAAAGAGAGAGAGAGAGATGTTAAGAAAGAGAGAGAGAGAGAGAGAGAG 2395
 QY 2090 CTGAAG 2149
 Db 2396 GATTAATCAACACTTAATGTTG-----CAGCAAGCGAAACAGAGAGAGAGAGAGAG 2443
 QY 2150 CCAACAG 2209
 Db 2444 GTCACTCCACACACACATTAATTCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2503

Accession	Version	Keywords	Source	Organism	Reference Authors Title	Journal Features	Gene	CDS
2210								
2504								
2270								
2564								
2330								
2624								
2390								
2684								
2450								
2741								
2510								
2801								
2570								
2861								
2630								
2921								
29								
PFAMSP1								
DEFINITION								
ACCESSION								
VERSION								
KEYWORDS								
SOURCE								
ORGANISM								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
FEATURES								
source								
gene								
CDS								

	mat_peptide	/partial. /gene="MSP1" 73..658 /gene="MSP1" /product="merozoite surface protein 1" /note="42 kDa fragment" /evidence=experimental 859..1200 /gene="MSP1" /product="merozoite surface protein 1" /note="19 kDa fragment" /evidence=experimental
	mat_peptide	
BASE COUNT	496 a 143 c 171 g 393 t	
ORIGIN		
Query Match	14.0% Score 691; DB 3; Length 1203;	
Best Local Similarity	73.4%; Pred. No. 5.4e-141;	
Matches 883; Conservative	0; Mismatches 320; Indels 0; Gaps 0;	
QY 3727	GTCATCATTTGGCTTACCTTCTGCGCAGAGAGAGAGACTACGATGACCCTGCCAGGTG	3786
Db 1	GTAATCATAGTACCTAATTATTTGAGAAATCCCAAGAACATTTATGATGATTGAGCAACAATA	60
QY 3787	GTCACCGGTGAGGCTGTCACCTCTCCGCTGATGTATTAATCTTGCTGCCAATAAGSAGAC	3846
Db 61	GTAACGAGGAAGAAGTAGTACCTCTCCGTAATGTAAACATACCTTTCTAAATTTGAAAAT	120
QY 3847	GAATACGAAGTGCCTATCTAATAACCTTGCGAGCGCTCTATAGCTCTCAAGAACAG	3906
Db 121	GAATATGAGTTTTATATTATTTAAACCTTTACGAGGTGTTATGAAAGTTTAAAAAACAA	180
QY 3907	CTGGAGATTAACGATGATGACCTTCATCTCCAACGGAAGGACATCTGCAACGCCGCTT	3966
Db 181	TTAGAAATTAACGTTATGACATTTATATGTATTAATGTTAAAGATATTTTAAATTCACGATTT	240
QY 3967	AATAAGAGAGAAAATTTGCAAAACCTCTTGGAGAGCGACTGATCCCTATPAAGACCTG	4026
Db 241	AATAAAGCTGAAAATTTCAAATAATSTTTAGATCGATTTATTTCCATPATPAAGATTTA	300
QY 4027	ACCTCTCTTAACCTAGTGTCTCAAGGACCCATPCAAGTTCTCTCAATPAAGAGAGGGAT	4086
Db 301	ACATCAAGTAAATTTATGTTGTCAMAAACATCAATAATTTCTTAATAAGAAAAAGAGAT	360
QY 4087	AAATTTCTGTAGTACACATATPTCAAGAGCTCCATCGACACCGATATCAATTTCCGT	4146
Db 361	AAATTTCTTAACGATTAATATATTTTAAGATTTCAATAGATAGCGATATAAATTTTGCA	420
QY 4147	AATGATGTGCTGGGTTATTACAGATCCTGACGCAAAAATPACAGCTTGACCTTGACTCT	4206
Db 421	AATGATGTCTGGTATATTATTAATAATATTATTCGAAAAATATAAATTCAGATTTAATTTCA	480
QY 4207	ATTAAAGTATATCAAGATVAGCAAGCGAGATVAGAAAATPTCTGCCCTTCTGAT	4266
Db 481	ATTAAAAATATATTCACGACACCAACAGTGAATAATGAGAAATACCTTCCTTTTAAAC	540
QY 4267	AACATCGAACCCCTGTACAGAAGCAAGTGAACCAAAATCGACTCTCTGTAATTCACCTG	4326
Db 541	AATATTTGAGCCTTATATTAATAAACAGTTATGATTAATAATGATTTATTTGTAATTCATTTA	600
QY 4327	GAGGCAAGGTCTCACTATATCTTACGAGAAAGACCATGTGGAAGTTAAATCAAGGAG	4386
Db 601	GAAAGCAAAAGTTCTTAATTTATACATATAGAAATCAAAACGTAGAAGTTTAAATTAAGAA	660
QY 4387	CTGAACCTACCTCAAAAACCAATCCAAACAGCAAGCTGAGATTTCAAGAAAATTAACAATTTTC	4446
Db 661	CTTAATTTACTTAAAAACAATTCACAGCAACAATTTGGCAGATTTTAAAAAAAATATACATTTTC	720
QY 4447	GTCGAATTTGACAGCTCTGTACCGATTATACCAACAACAATCTCTCTGACCAAGTTTCTG	4506
Db 721	GTTGGAATTTGCGATTTATTCACAAAGATATTAACCTATATTAACCTTATTCACAAAGTTTCTT	780
QY 4507	TTCACCTGGCATGTGTGTGAAAACCTGCCCAAAACAGTGTGAGGAATCTGTGTGACGGC	4566

Db 781 AGTACAGTATGGTTTGTAAATCTTGCTAAACCCTTTTATCTAATTACTTGATGA 840

Oy 4567 AACCTGCAGGGCATGCTGAACATCTCCAGCAACCAATGGGTGAAGAAACAGCCCCAG 4626

Db 841 AACTTGCAGAGTATGTTAAACATTTTCACACACCAATGGTAAACAAATGTCCACAA 900

Oy 4627 AATAGCGGCTGTTTTCAGGCACTGACGAGCGCGAAGAGTGCATGTCCTCGAAGTAC 4686

Db 901 AATTCTGATGTTTTCAGACATTTAGATGAAGAAGAAATGTAATGTTTATTAATTAC 960

Oy 4687 AAACAAGAAGATAGTACGCTGGAGAACCAACCCCTACTGCAATGAAACAAATGCG 4746

Db 961 AAACAAGAAGTATGTAATGTTTGAATAATCCAAATCTTACTGTAACCAATAATAGT 1020

Oy 4747 GGGTGTAGCGGATGCTTAATATGACCGGAGAAACAGGCGCTTAACGGAAGAAATC 4806

Db 1021 GGATGTGATGCAAGATGCCAAATGTACCGAAGAAATGATCAGTAGCAAGGAAATAC 1080

Oy 4807 ACATGCGAGTGTACTAAGCCGACCTCCATCCATCTTTCAGCGAGATTTTGTCTCAGC 4866

Db 1081 ACATGGAATGTACTAAGCACTGATTTCTTATCCACTTTTGTGATGTTTTCGCAATTGC 1140

Oy 4867 TCTAATTTCTCGGCACTCTCTCTGCTGATCTCTCATCTGATCTGTACAGCTTCATC 4926

Db 1141 TCTAATCTTGTAGATATCATCTTATTAATACTCATGTTAATATATACAGTTTCATT 1200

Oy 4927 TAA 4929

Db 1201 TAA 1203

RESULT 30

AF325919 1131 bp mRNA linear INV 13-FEB-2001

LOCUS AF325919

DEFINITION Plasmidium falciparum merozoite surface protein 1 (MSP-1) mRNA, partial cds.

ACCESSION AF325919

VERSION AF325919.1 GI:12751398

KEYWORDS

SOURCE

ORGANISM

Plasmidium falciparum.

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 1131)

L1,X.R., Goel,V.K., Liu,S.C., Chisholm,A.H. and Oh,S.S.

42 kDa subfragment of MSP-1 gene of Plasmodium falciparum FCB-1 isolate

JOURNAL

REFERENCE

2 (bases 1 to 1131)

L1,X.R., Goel,V.K., Liu,S.C., Chisholm,A.H. and Oh,S.S.

Direct Submission

Submitted (04-DEC-2000) Section of Hematology/Oncology Research, St. Elizabeth's Medical Center, Tufts University School of Medicine, 736 Cambridge Street, Boston, MA 02135, USA

JOURNAL

TITLE

Location/Qualifiers

1. 1131

/organism="Plasmodium falciparum"

/isolate="FCB-1"

/db_xref="taxon:5833"

/country="Colombia"

<1. 1131

/gene="MSP-1"

<1. 1131

/gene="MSP-1"

/codon_start=1

/product="merozoite surface protein 1"

/protein_id="AAK07641.1"

/db_xref="GI:12751399"

/translation="AVTPSYDNLISLIENEYLYLKLPLAGVRSLLKOLENNVMT NVNKKDILNSRENNKLVESDLIPYKDLTSNVVVKDPEENKREKDKLSSY NYIKSDIDTIDNFANDVLYGKILSEKKYSDLSIKKYINDKQENKEKYLPLNNIE LYKTVNKKIDLPVHLERKVLNTYEKSNEVKIKELVLTATIDDKLADFKKNPFV IADLSTYDNNHNLTKFLSTGMEFNLAFTVLSMLDGNLQGLNLTISDHCKVKKQCPQ NSGCFRIIDREBECKLLANYKQEGDKCVENPNPCNENNGCDADACETBDSGNGK

BASE COUNT 467 a 136 c 154 g 374 t

ORIGIN

Query Match 13.3%; Score 655.8; DB 3; Length 1131;

Best Local Similarity 73.7%; Pred. No. 3e-133;

Matches 834; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

Oy 3799 GCTGTACATCTCTCCGTGATGATTAACATTTCTGCCAAATCGAAGACGATGCAAGTG 3858

Db 1 GCAGTAACCTCTCCGTAATGATTAACATCTTCTTAAATGCAATGATGTGAGTT 60

Oy 3859 CTTATCTGAAACCTCTGGAGCGCTTATGCTCTCAAGAAAGAGTGGATTAAC 3918

Db 61 TTATATTTAAACCTTTACGAGGTGTTTATAGAGTTTAAACAAATAGAAATTAAC 120

Oy 3919 GTGATGACCTTCAATGTCAACGTAAGACATTTCTGACAGCGCTTTAATAGAGAA 3978

Db 121 GTTATGACATTTATATGTTAATGTTAAGATATTTTAAATTCACGATTTAATACGTGA 180

Oy 3979 AATTTCAGAGCTCTTGAGAGCGACTTGATTTCCCTATTAAGACCTGACCTCTTAAC 4038

Db 181 AATTTCAGAGCTCTTGAGAGCGACTTGATTTCCCTATTAAGACCTGACCTCTTAAC 240

Oy 4039 TACGTTGTCAAGACCATTAACATTAAGTTCCTCAATTAAGAGAGGATTAATTCGTCT 4098

Db 241 TATGTTGTCAAGACCATTAACATTAAGTTCCTCAATTAAGAGAGGATTAATTCGTCT 300

Oy 4099 AGTTACACTATATCAAGACCTCCATCGACACCGATATCAATTCGCTATGATGTCTG 4158

Db 301 AGTTATATTAATTAATTAAGATTCATTAAGATTAAGATTAATTAATTCGCAATGATGTTCT 360

Oy 4159 GGGTATTACAGATCTCTGAGAGCGCAAAATACAGCTGACCTTGACTTTTAAAGATAT 4218

Db 361 GGATATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420

Oy 4219 ATCAACGATTAAGAGCGAGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 4278

Db 421 ATCAACGATTAAGAGCGAGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 480

Oy 4279 CTGTACAGACAGTGAACGACAAATCGACTTCTGTAATTCACCTGAGGCGCAAGTTC 4338

Db 481 TTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540

Oy 4339 CTCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4398

Db 541 CTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 600

Oy 4399 AAACCAATCCAGACAGCTGCGAGATTCAGAGAAATTAATTAATTAATTAATTAATTAAT 4458

Db 601 AAACCAATCCAGACAGCTGCGAGATTCAGAGAAATTAATTAATTAATTAATTAATTAAT 660

Oy 4459 GACCTGTACAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4518

Db 661 GATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720

Oy 4519 GTGTTCGAAACCTGCGCAAAACAGTCTGACCAATTCGCTGAGGCAACCTGCAAGGCG 4578

Db 721 GTTTTGAATAATCTGTAAACCGTTTATCTTATTAATTAATTAATTAATTAATTAATTAAT 780

Oy 4579 ATGCTGACATCTCCAGACCAATGCGGTAAGAAACAGTGGCCCGCAGAAATGAGCGCTGT 4638

Db 781 ATGTTAAACATTTTCACACACCAATGCGTAAACCAATGCTCCACAAATTTCTGATGT 840

Oy 4639 TTCAGCATCTGAGAGCGCGAAGAGTGAAGTGTCTCTCACTACATCAACAAAGAGTA 4698

Db 841 TTCAGCATCTGAGAGCGCGAAGAGTGAAGTGTCTCTCACTACATCAACAAAGAGT 900

Oy 4699 GATTAAGTGTGAGAGAGCAACCAATCTTACCTGATTAAGAAACAAATGAGGAGTGTACGCC 4758

Db 901 GATTAATGTGTTGAATAATCTTACCTGATTAAGAAACAAATGAGGAGTGTATATCA 960

Oy 4759 GATGCTAATGACAGGAGAAAGACAGCGGCTCTAACGGAAGAAATACATGCGAGTGT 4818


```

Db 961 GATGCAAAATGTACCAAGAGATTCAGTAGCAGCAAGGAAAGAAATACATGTGATGT 1020
Qy 4819 ACTAAGCCGACACTTCATCCAGCTCTGACGAGGATTTTTCGCCAGCTCATTTTCCTG 4878
Db 1021 ACTAAGCCGATTTTATTCACATTTTCGATGATTTTTCGAGTTCTCCTCACTTCTTA 1080
Qy 4879 GGCATCTCTCTCTGCTGATCTCATGCTGATCTGATCTGATCTGATCTGATCTGAT 4929
Db 1081 GGAATATCATTTCTTATTAATACATGTTAATATTAATACAGTTTCAATTNA 1131

RESULT 31
PFASURFPRO 1065 bp DNA linear INV 26-JUL-1993
LOCUS Plasmodium falciparum merozoite surface protein (MSP-1) gene
DEFINITION
sequence.
ACCESSION L20092
VERSION L20092.1 GI:309745
KEYWORDS merozoite surface protein.
SOURCE Plasmodium falciparum (strain Vietnam Oak Knoll) blood stage DNA.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1065)
AUTHORS Louis-Willeman, V., Shi, Y., Collins, W. and LaI, A.
TITLE Primary amino acid sequence of the carboxyl-terminal region of the
moroite surface protein(MSP-1) of Plasmodium falciparum Vietnam
Oak knoll (FVO) strain
JOURNAL Unpublished (1993)
FEATURES
source location/Qualifiers
1..1065
/organism="Plasmodium falciparum"
/strain="Vietnam Oak Knoll"
/db_xref="taxon:5833"
/dev_stage="blood stage"
BASE COUNT 446 a 125 c 150 g 344 t
ORIGIN
Query Match 12.6%; Score 623.6; DB 3; Length 1065;
Best Local Similarity 74.2%; Pred. No. 3.5e-126;
Matches 788; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

Qy 3799 GCTGTACACCTTCCTCGATGTATTAATCTCTGCCAAATTCAGAAACGATACGAAGT 3858
Db 1 GCGGTACTCTCTCCGTAATGATTAACATCTTCTTAAATTTGAATGAAATGAGCTT 60
Qy 3859 CTCTATCTGAAACCTCTGGACGCTCTATAGCTCTCTCAAGAAACAGCTGGAATATAC 3918
Db 61 TTATATTAAACCTTAGAGAGCTTTATAGAAGTTTAAAGAAACATTTAGAAATATAC 120
Qy 3919 GTGATGACCTTCAATGTCAACGTGAAGACATTTCTGAACAGCCGCTTTAATAAGAGAA 3978
Db 121 GTATATGACATTTAATGTATTAAGTAATTTTAAATTCACGATTTAATAACGGAA 180
Qy 3979 AATTCAAGAACTCTTGGAGAGCAGCTGATTCCTATTAAGACCTGACCTCTCTAAC 4038
Db 181 AATTCAAAATTTTGTAGATCATGATTTAATTCATTAATAATTAACATCAAGTAAT 240
Qy 4039 TACGTGTCAAGAGACCATTCAGTTCTCCATTAAGAGAGAGGATTAATTTCTGTCT 4098
Db 241 TATGTTGTCAAGATCCATTAATTTCTTAATAAGAAAGAGATTAATTTCTTAAGC 300
Qy 4099 AGTTAACAATATCAAGACTCCATCGACAGCATATCAATTTCCGTATGATGTCTG 4158
Db 301 AGTTAATAATTAATTAAGATTCATTAAGATTAATTAATTTTTCGAATATGATGTTCT 360
Qy 4159 GGGTATTACAAGATCTCTGACGAAATATCAAGCTGACCTGACTCTATTAAGATAT 4218
Db 361 GGAATATTAATAATTAATTCGAAATATTAATAATTAATTTAGATTCATTAATAAATAT 420
Qy 4219 ATCAAGATTAAGAGAGGAGATTAAGAAATATCTGCTCTCGAATAATACATCGAAAC 4278
Db 421 ATCAAGAGCAAAACAGGTGAATATGAAATATACCTCTTTTAAACAATATTGAGACC 480

```

```

Qy 4279 CTGTACAGACACTGGAACGCAAAATGACCTCTCTGTAATTCACCTGAGCCAAAGTC 4338
Db 481 TTATATTAAGAAACAGTTATATGATTAATTAATTTGATTAATTCATTTAGAGAAAGT 540
Qy 4339 CTCAACTATTAATTAAGAGAGCAATGAGAGTTAAATTAATTAAGAGCTGACCTCT 4398
Db 541 CTAAATTAATTAATTAAGAGAGCAATGAGAGTTAAATTAATTAAGAGCTGACCTCT 600
Qy 4399 AAACATCCACAGCAAGCTGAGAGATTTCAAGAAATTAATTAATTAATTCGAGAAATGCA 4458
Db 601 AAACATTTCAAGCAAAATTTGAGAGATTTTAAATTAATTAATTAATTCGAGAAATTCGT 660
Qy 4459 GACCTGTCTACCGATTAATTAACCAACAAATCTCTGACCAAGTTCTGTCTGACCTGATG 4518
Db 661 GATTTATCAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720
Qy 4519 GTGTTGAAACACTGCGCAAAACAGTGTGAGCAATCTGTGACAGGCAACCTGCAAGGC 4578
Db 721 GTTTTGAATATCTTCTTAACCGTTTATTAATTAATTAATTAATTAATTAATTAATTAAT 780
Qy 4579 ATCTGACATCTCTCCAGCACCAGTCTGTAAGAAACAGTGTGACAGATTAATTAATTAAT 4638
Db 781 ATGTTAATCAATTTTCAACAGCAATGCTTAATTAATTAATTAATTAATTAATTAATTAAT 840
Qy 4639 TTCAAGCATCTGACAGAGCCGAGAGAGTGTCTCTGACTCACTCAACAAAGAGCA 4698
Db 841 TTCAAGCATTTAGATGAAAGAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 900
Qy 4699 GATTAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4758
Db 901 GATTAATGTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
Qy 4759 GATGCTAATATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4818
Db 961 GATGCAAAATGTACCAAGAGATTCAGTAGCAGCAAGGAAAGAAATACATGTGATGT 1020
Qy 4819 ACTAAGCCGACACTTCATCCAGCTCTGACGAGGATTTTTCG 4860
Db 1021 ACTAAGCCGATTTTATTCACATTTTCGATGATTTTTCG 1062

RESULT 32
PFMSA1AC 999 bp DNA linear INV 22-MAY-2002
LOCUS Plasmodium falciparum clone 808 major merozoite surface antigen
DEFINITION (p190) gene, partial cds.
ACCESSION M77715
VERSION M77715.2 GI:21040168
KEYWORDS T-cell epitope, major merozoite surface antigen, tripeptide repeat.
SOURCE Plasmodium falciparum.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 999)
AUTHORS Jongwutives, S., Tanabe, K., Nakazawa, S., Yanagi, T. and Kanbara, H.
TITLE Sequence variation in the tripeptide repeats and T cell epitopes in
p190 (MSP-1) of Plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
COMMENT On May 21, 2002 this sequence version replaced gi:309698.
FEATURES
source location/Qualifiers
1..999
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/map="9"
/clone="808"
/sex="asexual"

```



```

gene
/dev_stage="erythrocytic"
<1..>999
/gene="p190"
<1..>999
/gene="p190"
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AA29662.2"
/db_xref="GI:21040169"
/translation="ALEDAVLITGYSLEFQEKWVLENGTSIAVTTSTPGSGSVASGG
SGSVASGSGSVASGSGSVASGSGNSRNPSSDSDAKSYADLKHVRNY
LITIKELYPOLFDLTNMLTCDNIHGFYKLIDYEINELLYKLNFYFDLIRAKLN
DVCANDYCOIPFNLIKIRANEDLVKLEIVGYRKPLDNKIDNVGKMEYIKRKTIE
INLEIEESKTIIDKRNKATREEEKKLYOAOYDLSIVKOLEEANHLSVLEKRTIDL
KKNENIKELLDKINIKRPPRANGNTPTNLLDKKRIEHEKEIKELAKTIKENID"
BASE COUNT      418 a      127 c      159 g      295 t
ORIGIN

Query Match      11.5%; Score 570.2; DB 3; Length 999;
Best Local Similarity 73.2%; Pred. No. 1.9e-114;
Matches 731; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

OY 109 GCTTTGAAGATCCGCTCTTACCGGATACAGCTGTTCCAGAGAGAAGATGCTGTG 168
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 1 GCTTTGAAGATGAGTATTTGACAGGTATAGTTATTTCAAAAGGAAAAATGTAATTA 60

OY 169 AATGAAGGACAGATGGCCGCGCTTACACCAGCACCCGGTTCTAAAGGCTGTG 228
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 61 AATGAAGGACAAAGTGAAGACAGCTGTACACTAGACACGCTTCAAAAGGTTTCAATT 120

OY 229 GCTACGCGTGGCTCCGCGGTCTGTGCGCTTGGGGGTTCCGTCGCGCTCCGGCGGAC 288
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 121 GCTTACAGGTGGTTCAGTGGCTCAGTGGCTTCAAGTGGCTCAGTGGCTCAGTGGCTCA 180

OY 289 GTGGCATCAGGTGCTCAGTGGCAAGCGGCGTCCGGAGACATCGAACAATCA 348
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 181 GTTGGCTCAGGTGGCTCAGTGGCTTCAAGTGAATTCAGACATCAATCAATCT 240

OY 349 TGTGACACTCTAGGATCCGACGCAAGTCTACGCGGACCTCAAGACGAGAGACA 408
      || || || || || || || || || || || || || || || || || || || || ||
DB 241 TCGATTAATTCAGATTCAGATTCGTAATCTTTCGCGATTTTAAACAGAGACAGA 300

OY 409 AACTATCTCCCTCATCATCAAGAGCTGAAGTACCCACATGTTGACCTCATATCAT 468
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 301 AATTACTGTTAATCAATCAAGACCAATATCCCACTCTTGAATTAATCAATCAT 360

OY 469 ATGCGACACTGTGATACATTCATGCTCAATATCTGATTCAGCGTTACGAGAG 528
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 361 ATGTTAATCTGTTGATATATATTCATGTTCAATATTTAATTCATGATATGAAGA 420

OY 529 ATCAATGAACCTCTGTACAAAGTGAATTTCTACTGCTGCTAAGGGCCAAAGTGAT 588
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 421 ATTAATGAATTTATATTAATTAATTAATTTTATTTGATTTTATAGACCAAAATTAAT 480

OY 589 GACGTTGGCCCAATGACTATTTGTCAAAATTCATTCATTTGAGATTCAGACCAAG 648
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 481 GATGATGCTAATGATATATTGCAAAATCTTCAATCTTAAATTTGTCGCAATATGA 540

OY 649 TTGGACGATTTGAAGAAGTGTGCTTCGATATCCGACGCTCTGACACATCAAGAG 708
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 541 TTAGACGACTTTAAAAAATGTTGCGATATAGAAAAACCATTAATATTTAAAGAT 600

OY 709 AATGGGGAAGATGAGATATATTAAAAAAGATTAAGACCATGAGAACATTTAAC 768
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 601 AATGATGAGAAAAATGAGACATTACATTAATAAAAAATTAAGAAAAATTAATTAAT 660

OY 769 GAGCTGATCGAAGATCCAAAAAGACATAGACAAAAATAGATGACCAAGAGAGAA 828
      || || || || || || || || || || || || || || || || || || || || ||
DB 661 GAATTAATTAAGAGAAATAGAAAAAATTAATTAAGAAATGACACTTAAGAGAGAA 720

OY 829 GAAAAAGAAAGTGTACAGGCCAGTACGACCTGCTCATCTATTAACAACAAGCTTGA 888
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 721 GAAAAAATAAATTTATACCAAGCTCAATATGATCTTTTATTATTAACAATTAAGAA 780

```

```

OY 889 GAAGCCCATACCTCATCAGCTGAGAGAGCCATAGACACCTCAAGAAATGA 948
      ||||| ||||| || || || || || || || || || || || || || || || ||
DB 781 GAAGCACAATTAATTTAATAAGCGTTTGAAGAAACGATTTGACCTTTAAAAAATGA 840

OY 949 AATATCAAGAACTGCTGCACAAAGATTAATGAATTAAGAAATCTCCGACCACTCT 1008
      || || || || || || || || || || || || || || || || || || || || ||
DB 841 AACATTAAGAAATTTACTTATATAAATGAATTAATTAATTAATTAATTAATTAAT 900

OY 1009 GGAACACCCCTTACACGCTGCTGACAGAACAGACATGAGAGACACGAGAAAG 1068
      || || || || || || || || || || || || || || || || || || || || ||
DB 901 GGAATAATCACCAAAATCTCTCTGATTAAGACAAAAAATGAGAGACGAAAAAGAA 960

OY 1069 ATCAAGAGATCGCCAAACCATTAAGTCAACATGAT 1107
      || || || || || || || || || || || || || || || || || || || || ||
DB 961 ATTAAGAAATTCGCAAAACTATTAATTAATTAATGAT 999

RESULT 33
PFMSA1AG
LOCUS      999 bp      DNA      linear      INV 22-MAY-2002
DEFINITION Plasmodium falciparum clone 827 major merozoite surface antigen
            (p190) gene, partial cds.
ACCESSION M77719
VERSION    M77719.2 GI:21040176
KEYWORDS   T-cell epitope; major merozoite surface antigen; tripeptide repeat.
SOURCE     Plasmodium falciparum.
            Plasmodium falciparum.
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 999)
            Jongwutives,S., Tanabe,K., Nakazawa,S., YanagI,T. and Kanbara,H.
            Sequence variation in the tripeptide repeats and T cell epitopes in
            p190 (MSA-1) of Plasmodium falciparum from field isolates
            Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
JOURNAL    MEDLINE
            PUBMED 1373473
REFERENCE  2 (bases 1 to 999)
            Jongwutives,S.
            Direct Submission
            Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn
            University, Rama4, Pathumwan, Bangkok 10330, Thailand
            On May 21, 2002 this sequence version replaced gi:309705.
FEATURES
            source
            1..999
            /organism="Plasmodium falciparum"
            /db_xref="taxon:5833"
            /map="9"
            /clone="827"
            /sex="asexual"
            /dev_stage="erythrocytic"
            <1..>999
            /gene="p190"
            <1..>999
            /gene="p190"
            /codon_start=1
            /product="major merozoite surface antigen"
            /protein_id="AA29662.2"
            /db_xref="GI:21040177"
            /translation="ALEDAVLITGYSLEFQEKWVLENGTSIAVTTSTPGSGSVASGG
            SGSVASGSGSVASGSGSVASGSGNSRNPSSDSDAKSYADLKHVRNY
            LITIKELYPOLFDLTNMLTCDNIHGFYKLIDYEINELLYKLNFYFDLIRAKLN
            DVCANDYCOIPFNLIKIRANEDLVKLEIVGYRKPLDNKIDNVGKMEYIKRKTIE
            INLEIEESKTIIDKRNKATREEEKKLYOAOYDLSIVKOLEEANHLSVLEKRTIDL
            KKNENIKELLDKINIKRPPRANGNTPTNLLDKKRIEHEKEIKELAKTIKENID"
BASE COUNT      418 a      127 c      159 g      295 t
ORIGIN

Query Match      11.5%; Score 570.2; DB 3; Length 999;
Best Local Similarity 73.2%; Pred. No. 1.9e-114;
Matches 731; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

OY 109 GCTTTGAAGATCCGCTCTTACCGGATACAGCTGTTCCAGAGAGAAGATGCTGTG 168
      ||||| ||||| || || || || || || || || || || || || || || || ||

```


Db 1 GCTTTAGAAAGATGACAGTATTGACAGGTTAATGTTATTTCAAAAAGAAAAATGATTA 60
 QY 169 AATGAAGGACGAGTGGGACGCGCTTACCAACGACACCGCGTCTTAAAGGCTGTG 228
 Db 61 AATGAAGAACAGTGGACAGCTGTTACAAAGTACAGTACAGTCTGTTCAAGGCTGTT 120
 QY 229 GCTAGCGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGGTCCGTCGCTCCGGCGAGC 288
 Db 121 GCTTACAGGTGGCTCAGGTGGCTCAGTGTCTTCAAGGTGGCTCAGTGTGGCTCA 180
 QY 289 GTGGCATCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCA 348
 Db 181 GTTGTCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCA 240
 QY 349 TCTGCAACTCTAGCGATTCGACGCGCAAGTCTCCAGCCGCTCAAGCAACGAGTGA 408
 Db 241 TCAGATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 409 AACTATCTCTCAGTATCAAGGAGCTGAAGTACCAGTGTTCGACCTCAGTATCAT 468
 Db 301 AATTACTTGTAACTATCAAGGAGCTGAAGTACCAGTGTTCGACCTCAGTATCAT 360
 QY 469 ATGCTGACACTGTGCTGATCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
 Db 361 ATGTTAACTTGTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
 QY 529 ATCAATGAACTCTGTAACAAGTGAATTTCTACTGCTGCTGCTGCTGCTGCTGCTGCT 588
 Db 421 ATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480
 QY 589 GACGTTGGCCGCAATGCTATTGCAAAATTCATTCATTTGAAGATCAGACCAACGAG 648
 Db 481 GATGTAAGTCTCAATGATTTGCAAAATTCATTCATTTGAAGATCAGACCAACGAG 540
 QY 649 TTGAGCTATTGGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 708
 Db 541 TTGAGCTATTGGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 QY 709 AATGTGGGAAAGATGGAAGATTATTTAAAGATTAAGAACCATCGAACATTAC 768
 Db 601 AATGTGGGAAAGATGGAAGATTATTTAAAGATTAAGAACCATCGAACATTAC 660
 QY 769 GACCTGATCAAGATTCACAAAAGACCATGACAAAATTAAGATTCACAAAAGACGAG 828
 Db 661 GAATTAATTAAGAACATGAGTGAAGAACATTAAGAACATTAAGAACATTAAGAAC 720
 QY 829 GAAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
 Db 721 GAAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
 QY 889 GAAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 948
 Db 781 GAAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 QY 949 AATATCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1008
 Db 841 AATATCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 QY 1009 GGAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1068
 Db 901 GGAAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 QY 1069 ATCAAGAGATCGCAAAAGACATTAAAGTCAACATAGAT 1107
 Db 961 ATCAAGAGATCGCAAAAGACATTAAATTAATTTGAT 999

RESULT 34
 PFAMSA1AM 999 bp DNA linear INV 22-MAY-2002
 LOCUS PFAMSA1AM
 DEFINITION Plasmodium falciparum clone 837 major merozoite surface antigen
 ACCESSION M77725

VERSION M77725.2 GI:21040188
 KEYWORDS t-cell epitope; major merozoite surface antigen; tripeptide repeat.
 SOURCE Plasmodium falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolates; Apicomplexa; Haemosporidia; Plasmodium.
 REFERENCE 1 (bases 1 to 999)
 AUTHORS Jongwutives,S., Tanabe,K., Nakazawa,S., Yanagi,T. and Kanbara,H.
 TITLE Sequence variation in the tripeptide repeats and T cell epitopes in
 P190 (MSA-1) of Plasmodium falciparum from field isolates
 JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
 MEDLINE 92228018
 PUBMED 1373473
 REFERENCE 2 (bases 1 to 999)
 AUTHORS Jongwutives,S.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn
 University, Rama4, Pathumwan, Bangkok 10330, Thailand
 COMMENT On May 21, 2002 this sequence version replaced gi:309716.
 FEATURES
 source
 1..999
 /location=/Qualifiers
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /map="9"
 /clone="837"
 /sex="asexual"
 /dev_stage="erythrocytic"
 <1..>999
 /gene="p190"
 <1..>999
 /gene="p190"
 /codon_start=1
 /product="major merozoite surface antigen"
 /protein_id="AAA29672.2"
 /db_xref="GI:21040189"
 /translation="ALEDAVLTGYSLFQEKMYLNSTGTAVTSTPSKGSVASG
 SGGSVASGGSVASGVSAGSVASGSGRSRTNSDSDASVDLKRVRNY
 LTIKIKYKPOLFDLNNHMTLDNDHNGKRYLLDGEELNELLYKTFEYDLRLATLN
 DVCANDYCOIPENLKTIRANEDLVKRLVGYRPLDNIQNGKMDYIKKKRTIEN
 IMELIESKRTIDKNRNATKEERKRLYQADLSLYNQLEAHNLISVLEKRTDTL
 KKNENIKELIDKINEIKNPSPANGTPTPLDLDNKKIEHEKEIKETIKFNID"
 BASE COUNT 418 a 127 c 159 g 295 t
 ORIGIN
 Query Match 11:58; Score 570.2; DB 3; Length 999;
 Best Local Similarity 73.2%; Pred. No. 1.9e-114;
 Matches 731; Conservative 0; Mismatch 268; Indels 0; Gaps 0;
 QY 109 GCTTTGGAAGATGCGGCTTACCGGATFACAGCTGTTCCAGAGAGAGATGCTGTG 168
 Db 1 GCTTTGGAAGATGCGGCTTACCGGATFACAGCTGTTATTTCAAAAAGAAAAATGATTA 60
 QY 169 AATGAAGGACGAGTGGGACGCGCTTACCAACGACACCGCGTCTTAAAGGCTGTG 228
 Db 61 AATGAAGAACAGTGGACAGCTGTTACAAAGTACAGTACAGTCTGTTCAAGGCTGAGTT 120
 QY 229 GCTAGCGGTGGCTCCGGTGGTCTGTGGCTCTGGGGGGTCCGTCGCTCCGGCGAGC 288
 Db 121 GCTTACAGGTGGCTCAGGTGGCTCAGTGTCTTCAAGGTGGCTCAGTGTGGCTCA 180
 QY 289 GTGGCATCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCA 348
 Db 181 GTTGTCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCAGGTGGCTCA 240
 QY 349 TCTGCAACTCTAGCGATTCGACGCGCAAGTCTCCAGCCGCTCAAGCAACGAGTGA 408
 Db 241 TCAGATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 409 AACTATCTCTCAGTATCAAGGAGCTGAAGTACCAGTGTTCGACCTCAGTATCAT 468
 Db 301 AATTACTTGTAACTATCAAGGAGCTGAAGTACCAGTGTTCGACCTCAGTATCAT 360
 QY 469 ATGCTGACACTGTGCTGATCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528

Db 361 ATGTTAACTTTGTGTGATTAATATTCATGCTTCAAAATATTAATGATGATGATGAAGA 420
QY 529 ATCAATGACCTCTGTACACTTGAATTTCTACTTCCACTTGTAGGCCCAACTGAAT 588
Db 421 ATTTATGATTAATTAATTAATTAACCTTTATTTTGAATTAAGCAAAATTAAT 480
QY 589 GACCTTTGGCCCAATGACTATGTCAAATTCATTTGGAATGAGACCAAGAG 648
Db 481 GATATGATGCTTAATGATTAATTTCAAAATACCTTTCAATCTTAAATTCGTCAATGA 540
QY 649 TTGACGATTTGAAGAAGTTGGCTCTGGATATGCAAGCCCTCGACACATCAAGAGAC 708
Db 541 TTAGACGACTTAATAAACTGTGTGCGATATGAAACCATTAAGACATTAATTAAGAT 600
QY 709 AATGTGGAAAGATGGAAGATTAATTAATAAGATAAGACCATCGAGACATTAAC 768
Db 601 AATGTAGAAAAGATGAAGATTAATTAATAAAATTAATAAACCATTAAGAAATTAAT 660
QY 769 GAGCTGATCGAAGATCCAAAAGACCATAGACAAAATAAGAAATGCAACCAAGAGAA 828
Db 661 GAATTAATGAGAAAGTAAGAAAACAATTTGATTAATAATGACACTTAAGAGAA 720
QY 829 GAAAAGAAAGATTGTACAGCCAGCTAGCACTGTCTATTAACAAACAGCTTGA 888
Db 721 GAAAAAATAATTAATTAACAGCTCAATATGATCTCTTATTAATAATTAAGAA 780
QY 889 GAAGCCCATACCTCATCAGCGCTAGTGAAGGCAATGACCCCTCAAGAAATGA 948
Db 781 GAACCATATTAATTAATTAAGGCTTTAGAAAACGTATGACCTTTAAATAAATGA 840
QY 949 AATATCAAGAACTGTCTGACAGATTAATGAATTAAGAAATTCCTCCGCCCAACTCT 1008
Db 841 AACTTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
QY 1009 GGGAAACCCCTTAACAGCTCTGAGACAGACAAAGATAGAGAGACAGAGAAAG 1068
Db 901 GGAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
QY 1069 ATCAAGAGATCGCCAAACCATTAAGTTCAACATTAAT 1107
Db 961 ATAAAGAAATTTGCCAAACCTTAATTAATTAATTAAT 999

RESULT 35
PFAMSA1AA
LOCUS Plasmodium falciparum clone 806 major merozoite surface antigen
DEFINITION (p190) gene, partial cds.
ACCESSION M77713.2 GI:21040164
VERSION T-cell epitope: major merozoite surface antigen; tripeptide repeat.
KEYWORDS Plasmodium falciparum.
SOURCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ORGANISM *Plasmodium falciparum*.
REFERENCE 1 (bases 1 to 999)
AUTHORS Jongwutives,S., Tanabe,K., Nakazawa,S., Yanagi,T. and Kanbara,H.
TITLE Sequence variation in the tripeptide repeats and T cell epitopes in p190 (MSP-1) of plasmodium falciparum from field isolates
JOURNAL Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
MEDLINE 9228018
PUBMED 1373473
REFERENCE 2 (bases 1 to 999)
AUTHORS Jongwutives,S.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn University, Rama4, Pathumwan, Bangkok 10330, Thailand
REMARK Sequence update by submitter
COMMENT On May 21, 2002 this sequence version replaced gi:309694.
FEATURES
location/Qualifiers
1..999
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/map="9"

Query Match 11.5% Score 568.6; DB 3; Length 999;
Best Local Similarity 73.18; Pred. No. 4.3e-114;
Matches 730; Conservative 0; Mismatches 289; Indels 0; Gaps 0;
ORIGIN
BASE COUNT 418 a 128 c 159 g 294 t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

Db	721	GA	AAAAAAAAA	TTT	TACCA	AGCTCA	ATTTG	ATCTTT	GTATTT	TACAT	TAA	CAATT	AGAA	TTAGAA	780
Qy	889	GA	AGCCCAT	TAC	CTAT	CAG	CGTACT	GGAA	AGCCAT	GTAC	ACCTCA	AGAA	GAT	TCAA	948
Db	781	GA	AGC	ACAT	AAAT	TTT	TAT	ATAG	CGCTTT	AGAAAA	CGAT	TGAC	ACTTT	AAAAAAAA	840
Qy	949	AAT	ATCA	AA	GA	ACT	CTCG	CA	CAAG	ATTA	TGA	AAAT	TAA	CAAT	1008
Db	841	AAC	ATT	AA	GGAA	TTC	TCT	AT	GTAT	TAA	GAAT	TAA	ATTA	AAAAAT	900
Qy	1009	GG	GA	C	AC	CC	CTTA	AC	GGT	GTGG	CA	GA	CA	GA	1068
Db	901	GG	AAAT	TAC	CCCA	AAAT	TCTCT	CT	GTAT	TAG	AA	CA	AAAA	AAAAAT	960
Qy	1069	AT	CA	AA	GA	GAT	CG	CA	AA	CA	CTTA	AGTT	CA	CA	1107
Db	961	AT	TAA	AA	GA	AAAT	TG	CC	AA	AA	CTAT	TAA	ATTT	TAT	999

RESULT	36			
LOCUS	PFMSA1AF			
DEFINITION	PFMSA1AF	999 bp	DNA	linear
ACCESSION	Plasmodium falciparum clone 822B major merozoite surface antigen (p190) gene, partial cds.			INV 22-MAY-2002
VERSION	M77718			
KEYWORDS	M77718.2 GI:21040174			
SOURCE	T-cell epitope; major merozoite surface antigen; tripeptide repeat.			
ORGANISM	Plasmodium falciparum.			
REFERENCE	Plasmodium falciparum.			
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
TITLE	1 (bases 1 to 999)			
JOURNAL	Jongwutiwises,S., Tanabe,K., Nakazawa,S., Yanagi,T. and Kanbara,H.			
PUBLISHED	Sequence variation in the tripeptide repeats and T cell epitopes in P190 (MSA-1) of Plasmodium falciparum from field isolates			
REFERENCE	Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)			
JOURNAL	92228018			
PUBLISHED	1373473			
REFERENCE	2 (bases 1 to 999)			
AUTHORS	Jongwutiwises,S.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-May-2002) Faculty of Medicine, Chulalongkorn University, Rama4, Pathumwan, Bangkok 10330, Thailand			
COMMENT	On May 21, 2002 this sequence version replaced by J0309703.			
FEATURES	Location/Qualifiers			
SOURCE	1..999			

```

/db_xref="taxon:5833"
/map="g"
/clone="822B"
/sex="asexual"
/dev_stage="erythrocytic"
<1..>999
/gene="p190"
<1..>999
/gene="p190"
/codon_start=1
/product="major merozoite surface antigen"
/protein_id="AA02965.2"
/db_xref="GI:21040175"
/translation="ALEDAVLVTGSLFEQEKRMVLNEGTSQAVTTPSPGSKGSVASGG
SGGSVASGGSVASGGSVASGGSVASGGSRTSPNSSDDAKSYADLKHRYRNN
LLTLEKLEKLPOLFLDTNMLFLDNIHGFKLIDCGEINELLYKLNFPDLRLAKLN
DYVCANDYCOIPFNLIKIRANEDLDYLLKLVFGKRPDLNIDKGVAKKEDIKKNNKITEE
INLEIEESKRTIDKNNKATKEEKKRLQAOYDUSTYNNKLEAHNLLSVLEKRTDIT
KNEIEKELSLDOKINEIKNPPANSQNTPTLLDNNKKIEEHEKEIKELAKTIKEMID

```

Query Match	11.5%;	Score 568.6;	DB 3;	Length 999;
Best Local Similarity	73.1%;	Pred. No. 4.3e-114;		
Matches 730; Conservative	0;	Mismatches 269;	Indels 0;	Gaps 0;

QY	109	GCCTTGGAAATGCGCCTTACCGGATACAGCTGTCCAGAAAGAAATGCTGCG	168
Db	1	GCCTTAGAAGATGCAGTATTGACAGGTTATAGTTTATTCCAAAGSAAAAATG	60
QY	169	AATGAAGGAGCAGATGCGCGCCGTTACACAGACACCCGGTCTTAAGGCTGTG	228
Db	61	AATGAAGAAACACGGGAAACAGCTGTTACACATAGTACACCTGGTTCCAAAGG	120
QY	229	GCTAGCGGTGCTCGCGGTGGTGTGCGCCCTCGGGGGTTCGTCGCTCGCGCG	288
Db	121	GCTTCAGGTGGTTAGGTGGCTCAGTTGCTTCAGGTGGCTCAETTCCTCAG	180
QY	289	GTGGCATCAGGTGGCTCAGTGGCAGCGCGGCTTCGGGAAACGTGGAAGAAC	348
Db	181	GTTCCTCAGGTGGCTCAGTTGCTTCAGGTGGCTTCAGTAAATTCAGAGCT	240
QY	349	TCTGCACACTTACGATTTCCAGCGCAATCCTACCGGACCTCAAGACCGAG	408
Db	241	TCAGTAAATTCAGTATTCAGATGCTAAATCTTCAGCTGATTTAAACACG	300
QY	409	AACATCTCCCTACATTCACAGAGCTGAATACCACAGTTGTTGCACCTCA	468
Db	301	AATTACTGTTAACTATTCACAAAGAACTCAATATCTCACTCTTGATTTA	360
QY	469	ATGCGACACTGTGATTAACATTCAATGCTTAAATATCTGATTTAGCGGTT	528
Db	361	ATGTTAACTGTTGTGATTAATTCATCAGTTTCAAATTTAATGATGATTA	420
QY	529	ATCANTGAACCTCGTACAGTGAATTTCTTACTTGACTTGCTGTAAGGCC	588
Db	421	ATTATGAATTAATTAATTAATTAATTAATTTGATTTAATTAAGACAAAT	480
QY	589	GACGTTTGGCCCATGACTATGTGCAATTCATTAATTTGAAGTACGACCA	648
Db	481	GATGATGTGCTAATGATTAATGTCAAATACCTTAAATTCGTCCAAATGA	540
QY	649	TTGACGATTTGAAGAGTGGTCTTCGTGATTCGCAAGCGCTCTCGACA	708
Db	541	TTAGCGTACTTAAAAAACTTGTTGCGATTTAGAAAAACATTAATTTAAG	600
QY	709	AATGCGGAAAGATGGAAGTTAATTTAAAAAATTAAGAACCATTCGAA	768
Db	601	AATGAGGAAAAAATGGAAGTATCAATTTAAAAAATTAACCAATTAATTA	660
QY	769	GAGCTGATCGAAGATCCAAAAAGACCATAGACAAAAATTAAGATGCA	828
Db	661	GAATTAATTTGAAGAAGTAGAAAAACATTTGATAAATAAGATCACTAA	720
QY	829	GAAAGAGAGTGTGACGAGCCCATAGACCTCCATCTATTAACAAGCTGA	888
Db	721	GAATAAAAAAATTAATTAACCAAGCTCAATTAATGATCTTCTATTACAT	780
QY	889	GAAAGCCATTAACCTATCAGGTACTGGAAGACGAGTAAAGACACCTCA	948
Db	781	GAAGACATTAATTAATTAAGGTTTGAAGAAACGTAATGACCTTTAA	840
QY	949	AATATCAAGAACGCTCGACAGATTAATGAATTAATTAATTAATTAAT	1008
Db	841	AACATTAAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTA	900
QY	1009	GGAAGACACCCCTAACAGCTGTGACAAACAAACAAAGATAGAGGAC	1068
Db	901	GGAATTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTA	960
QY	1069	ATCAAGAGATGCCAAACCATTAATTAATTAATTAATTAATTAATTA	1107
Db	961	ATTAAGAAATTTGCCAAACATTAATTAATTAATTAATTAATTAATTA	999

RESULT 37				
AF061143				
LOCUS	AF061143	1133 bp	DNA	linear
DEFINITION	Plasmodium falciparum isolate HN2 meozoite surface protein 1			

(msp1) gene, partial cds.
 AF061143
 AF061143.1 GI:3821978
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Plasmodium falciparum.
 Plasmodium falciparum
 Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 1133)
 REFERENCE
 JIANG, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and
 Pluschke, G.
 Sequence diversity of the merozoite surface protein 1 of Plasmodium
 falciparum in clinical isolates from the Kilombero District,
 Tanzania
 Acta Trop. 74 (1), 51-61 (2000)
 JOURNAL
 MEDLINE
 20106724
 10643908
 PUBMED
 2 (bases 1 to 1133)
 REFERENCE
 JIANG, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
 Direct Submission
 Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical
 Institute, Socinstrasse 57, Basel 4002, Switzerland
 JOURNAL
 TITLE
 FEATURES
 location/Qualifiers
 1. .1133
 /organism="Plasmodium falciparum"
 /isolate="HN2"
 /db_xref="taxon:5833"
 /chromosome="9"
 /note="wild isolate"
 66. .>1133
 /gene="msp1"
 66. .>1133
 /gene="msp1"
 /product="merozoite surface protein 1"
 66. .>1133
 /gene="msp1"
 /note="p190; surface protein with spectrin binding
 activity"
 /codon-start=1
 /product="merozoite surface protein 1"
 /protein_id="AAC69742.1"
 /db_xref="GI:3821978"
 /translation="MKTIIFLCSFLPIITOCYTHESYELKKEALEDAVITGYS
 LFQEKVNLNKGSTGTAIVTSTPSKSGSVASGSGSVASGSAVS
 GSGNSRNTSPDSDAKSYADLKHRYQNYLFTIKELPELPDLTHMLCDN
 IHGFKYLIDYEINELLYKLNFYDLRAKLDACNSYQLEPFIKIRANEDLYK
 KIVEGRKPLIDNIDKDYKMDYIKKKTLIANINELIESKRTIDONKANDEGKK
 KLYOAOYNLFYIKLOLOEALNLSVLEKRLDITLKKNNIKKLDKINEIKNPANSG
 NTPNTLIDKKKIEHEEKEIKE"
 BASE COUNT 460 a 142 c 167 g 364 t
 ORIGIN
 Query Match 11 3%; Score 556.6; DB 3; Length 1133;
 Best Local Similarity 70.1%; Pred. No. 1.9e-111;
 Matches 748; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

Db 306 TCAGTGGTTCAGGTGAGTCTCAGTTCCTCAGTGGCTCAGTTGCTTCACTT 365
 Qy 310 GCAGCGGGGCTCCGGGACAGTCGANGAACCATCATCTGACCACTAGGATGCC 369
 Db 366 GCTTCAGGTGGTCAAGTAATTCAGACGATCAAAATCCCTCAGTAATTAAGTATCA 425
 Qy 370 GACGCCAAGTCTCTGCGGACCTCAAGCAGCAGTGAAGAACTATTCCTCATTCAAG 429
 Db 426 GATGCTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTTACTTCTCATTAA 485
 Qy 430 GAGCTGAAGTACCCACAGTGTTCGACCTCACTAATCATATGACCTGTGATNAC 489
 Db 486 GAATCAATATCCCGAATCTTGTATTAACATCATATGATTAACCTTGTGATNAT 545
 Qy 490 APTCATGCTCAATATATCATGATGAGCTTACGAGCAAGATCATGACCTCGTACAG 549
 Db 546 ATTCATGCTTCAATATATTAATGATGATGATGATGATGATGATGATGATGATGAT 605
 Qy 550 TTGAATTTCTACTTCGACTTCGACCTCAAGGCGCAAGCTGATGATGATGATGATGAT 609
 Db 606 TTAAACTTTTATATGATTTATTAAGACCAATTAATGATGATGATGATGATGATGAT 665
 Qy 610 TGTCAAATTCATTCATTTGAAGATCAGACCAAGCTGACGATTTGAAGAATTTG 669
 Db 666 TGTCAAATATCCCTCAATCTTAATATGCTGCAATATGATGATGATGATGATGATGAT 725
 Qy 670 GTCCTCGATATTCGACACCTCTGACAACTCAAGCAAGCATGTCGGAAGATGGAAGAT 729
 Db 726 GTTTGGATATGAAACCATGACCATATTAAGATATGATGATGATGATGATGATGATGAT 785
 Qy 730 TATATTAAGAAATTAAGACCATGACCATGATTAAGCATGATGATGATGATGATGATGAT 789
 Db 786 TACATTAAGAAATTAAGACCATGACCATGATTAAGCATGATGATGATGATGATGATGAT 845
 Qy 790 AAGACCATAGACAAATTAAGATGCAACCAAGGAGCAAGAAAGAAATGTTGACAG 849
 Db 846 AAACCAATGATCAAAATTAAGATGCAACCAAGGAGCAAGAAAGAAATGTTGACCA 905
 Qy 850 GCCCGATGACACCTGTCATCATTAACCAAGCTTGAAGACCATGATGATGATGATGATGAT 909
 Db 906 GCTCATATTAATCTTTTATTTATTAACATTAACATTAACATTAACATTAATTAATGAC 965
 Qy 910 GTACTGGAAGCGCATGACACACCTCAAGAAATGAATATCAAGAACTGCTGAC 969
 Db 966 GTTTAGAAACGATATGACACTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1025
 Qy 970 AAGATTAATGAATTAAGATTCCTCCGACCACTGTCGGAACACCCCTAACACGCTG 1029
 Db 1026 AAGATTAATGAATTAAGATTCCTCCGACCACTGTCGGAACACCCCTAACACCTG 1085
 Qy 1030 CTGGACAAGACAAAGATAGAGGACGAGAAAGATCAAGA 1076
 Db 1086 CTTGATTAAGACAAATAATGAGGACGACGACGACGACGACGACGACGACGACGAC 1132

RESULT 38
 AF286876
 LOCUS 1080 bp mRNA linear INV 25-JUL-2000
 DEFINITION Plasmodium falciparum clone FCB-1 merozoite surface protein 1
 (msp-1) mRNA, partial cds.
 ACCESSION AF286876
 VERSION AF286876.1 GI:9438194
 KEYWORDS
 SOURCE Plasmodium falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 1080)
 REFERENCE Goel, Y.K., Liu, S.-C., Chisth, A.H. and Oh, S.S.
 38 kDa subfragment of MSP-1 gene of FCB-1 strain of Plasmodium
 falciparum
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1080)
 AUTHORS Goel, Y.K., Liu, S.-C., Chisth, A.H. and Oh, S.S.

/chromosome="9"
/dev_stage="blood stage"
1. .1956
/gene="gpi190, MSA1, MSP1, PMMSA"
<1. .1956
/gene="gpi190, MSA1, MSP1, PMMSA"
/product="start=1
/product="merozoite surface antigen 1 gene, 3+-end"
/protein_id="CA84558.1"
/db_xref="GI:535258"
/db_xref="SPTREMBL:O25924"
/translation="FNRRKELGDKRMQIKLTLLEQLESKLSLNPNVYLNQSVF
FNRRKELIETENTLENTLILKHYGLVKKYNGESSPLKLESEVSIQEDNANKE
KPRALSKIDKLNIDNHLGKRLSKLSLSSGHLITELKVNKNNTYSPENNKVY
NEALSKYENLPEAKVTYVTPPOPDVTPSPSVRSVSGSTKEETQIPSSGLTE
LOOVLOANYDEDDSLVYLPIGSESDNEDYLDVYTGKATISYTMNDNISGPENEYD
VYLKPLAPYRSRKQIEKNITFTNINLNDILNSRLKRYFLDYESIDLMOKRHS
SNEYIILEDSTKLNSEQNTLNSYKIKESVENDIKFADGSLSYERKVLAKYKDLLE
SIRKVIKEEKEPSPPTPPSPAKTDEQKESKFLPFLNIETLYNNLVKIDYLI
NLKAKINDKNEKDEAHVKITKSLDKAIDKIDLFNHNDFEAIKRLINDTQDKML
GLISLGLVONFPTIISKLEGFODMLNISQOCYKQCCPOMSGCFRHLDERECK
CLINYQEDKCVENPMPNENNGGCDAAKCTEEDSGSNKRTTCECTKPDSPLEF
DGIFCSSNLFGLISFLILMLILYSFT"

BASE COUNT 855 a 263 c 265 g 573 t
ORIGIN

Query Match 11.1%; Score 546.4; DB 3; Length 1956;
Best Local Similarity 57.8%; Pred. No. 3.5e-109;
Matches 1134; Conservative 0; Mismatches 726; Indels 102; Gaps 5;

QY 3064 TTCGACAGAGAAGAGACAGTCGCGAGATATAGATCGATCAAGAACTGCTGCTC 3123
DB 1 TTTATATAGAAAAAAGAACTGCGCAGACAAATGCAATTAATAAACTTACTTTATA 60
QY 3124 AAGAGACAGCTTGAAGCAACTCACTCACTGAACAATCCGAAACAGCTACTGACAG 3183
DB 61 AAAAAACAATTAAGATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120
QY 3184 TTTCAGTGTCTTCAACAG 3243
DB 121 TTTTCTGTTTCTTTTAAACAAAAAAGAACTGAATAGCAAACTGAAACACATTA 180
QY 3244 GAGAAGACCAAGATTTCTTCACAACTACAAAGGCTCTGCAAGTATTTATATGCGAG 3303
DB 181 GAAAAACAAAAATATTATTGAACATTAATAGAGACTTGTAAATATTATTAATGTTGTA 240
QY 3304 TCTTCTCTCTGAGAGACTCTCCGAGAGAGCATCCAGACCGAGATACTAGCCAGC 3363
DB 241 TCATCTCCATTAATAAACTTTAAGTGAAGTATCAATTCACAGAGAAATTAATGCCAAT 300
QY 3364 CTGAGAACCTTCAAGTCCTGCTAGCTGAGTGAAGCAAGTGAAGCAACCTGACCTG 3423
DB 301 TTGAAAAAATTTAGAGCATTTAGTAATAATGATGAAAACTCAATGATTAATTTACATTTA 360
QY 3424 GAGAAGAGAGAGCTCAGCTACCTCTTACGCGAGCTGATCACTGATCGCGAGCTCAAG 3483
DB 361 GGAAGAAAAAATATTCTTTCTTATCAAGTGAATACATCTTTAATTAGCAATTAATAA 420
QY 3484 GAAGTCATTAGAACAAAGAACTACACCGGCATATGCCAGCGAGATTAATACAGAGCTG 3543
DB 421 GAATTAATTAATAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATAAAAGTT 480
QY 3544 AATAAGCACTGAGATCTTACAGAGAGTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 3596
DB 481 AACGAACTTTAAATCTTACGAAAAATTTTCTCCAGAGAGCAAAAGTTACAAAGTTGTA 540
QY 3597 -----CACTGTGTCTGAAATCTGCTCCGACACACTGGA 3632
DB 541 ACTCACCTCAACAGATGTAATCTCAATTCATTTATCTGTAAGGTAAAGTGAAGTGA 600
QY 3633 GCATCTCAACCTTAAGAA-----GCCTGATCTACTCATGTGCGAGCGAGATCCAAAT 3684
DB 601 GGATCCACAAAGAAAGAACAAATATCAACTTCAGGCTCTTTATTAAACAGAAATTACAA 660

QY 3685 ACAATTACCACTCAGACAGTGCAGATGAGTGCATGAGTGCATGATGCTGCTATC 3744
DB 661 CAACTAGTACATTTTCAAAATTTATGACCAAGAGATGATGCTTCTTATGTTATACCAAT 720
QY 3745 TTGCGCGAGAGCGAGAGAGACTAGATGACCTGCGCGAGTGTGACCGGTGAGGCTGTC 3804
DB 721 TTGAGAGATCCGAAGATATGAGCAATATTATGATCAAGTATGATCAAGTATGAGCAATTA 780
QY 3805 ACTCTTCCGATGATTGATTAATCTGTGCAAAATCGAAGCAATACGAAGTGTCTAT 3864
DB 781 TCTGCACAAAG---GATTAATATCTCTCAGAGATTTGAAAAATGAAATGATGATTAATAT 837
QY 3865 CTGAACCTCTGCGAGGCTGTATGCTCTCAAGAAACAGCTGAGATTAAGCTGATG 3924
DB 838 TTAACACCTTTAGCTGGGATATATGAGCTTTAAAAAACAAATTTGAAAAAACCTTTT 897
QY 3925 ACCTTCAATGTCACAGTGAAGACATTCGTGAACAGCCGCTTTAATAGAGAAAAATTC 3984
DB 898 ACATTTAATTTAATTTGACGATATCTTAATTCACGCTTGAAGAAAGAAAAATATTC 957
QY 3985 AAGAACGCTTGGAGAGAGCTGATTCCTATTAAGACCTGACCTCTTAACAGCT 4044
DB 958 TTGATGATTAAGATTCGATTTAATGCAATTTAAACATATATCTCTCAAAATGAATCAAT 1017
QY 4045 GTCAAGACCCATCAAGTCTCTCAATTAAGAGAGAGAGATTAATTTCTGCTAGTATC 4104
DB 1018 ATTGAGATTTATTTAATTTATGATTCAGACAAAAAACACACTTTTAAAAAGTTAC 1077
QY 4105 AACTATATCAAGAGACTCCATGACACCGATATCAATTTGCTAATGATGTCTGGGAT 4164
DB 1078 AAATATTAATAAGATCGTGAATAATGATATTAATTTGACAGAGAGATTAATGAT 1137
QY 4165 TACAAGATCTGAGCGAAAAATACAGTCTGACCTTGAATTTAAAAAGTATATCAAC 4224
DB 1138 TATGAAAAAGTTTGAAGCAATTAAGATGATTAATCAATTAATAAAAGTTATCAA 1197
QY 4225 GATAAG----- 4230
DB 1198 GAAGAAAAAGAGTCCCATCATCAACCAACCAACCTCCGTCAACGCAAAAAAGAC 1257
QY 4231 ---CAAGCGAAGTGAATAATATCTGCTCTGTAATTAATCAAGAACCTGTACAG 4287
DB 1258 GAACAAAAAGAAAGTAAGTCTCTCTTAAACCAATTAAGACCTTATACAT 1317
QY 4288 ACAGTGAACGCAAAATGACCTCTGTAATTAACCTGAGGCGCAAGTCTCAACAT 4347
DB 1318 AACTTAGTTAATAAATTTGACGATTAATTAATTAACCTTAAGGCAAAAGATTAACGATGT 1377
QY 4348 ACTTAGGAGAGACCAATGTGAAGTTAAATCAAGGAGCTGAACCTACCTCAAAACATC 4407
DB 1378 AATGTTGAAAAAGATGAAGCACTGTTAAATTAACCTTAATGATTTAAAGCAAT 1437
QY 4408 CAAGCAAGCTGCGAGATTTCAAGAAAAATTAACATTTCTGCGAATTTGACAGCTGTCT 4467
DB 1438 GATGCAAAATTAAGATCTTTTAAAAACCATTAACGCTCGAAGCAATTAATAAAATTTGATA 1497
QY 4468 ACCGATTTAACCACACAAATCTCTGACCAAGTTTCTGTCACATGCGATGTGTTGAA 4527
DB 1498 AATGATGATGAGAAAAAAGATGCTGCGCAAAATTAATTAAGAGGATTTG---TTCAA 1554
QY 4528 AACCTGCGCAAAAGCTGAGCAATCTGCTGCGACGCACTGACGAGGAGCTGTAAC 4587
DB 1555 AATTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1614
QY 4588 ATCTCCAGCAACATGCGTGAAGAAACAGTGCACCAAGATTAAGCGCTGTTGAGCAT 4647
DB 1615 ATTTCACACACCAATGCTTAATAAAACAAATGCTCCACAAATTTCTGATGTTTACAGAT 1674
QY 4648 CTGACAGAGCGGAGAGATGTAATGCTCTGAACTTCAACAAAGAAAGAGATTAAGTGC 4707
DB 1675 TTGATGTAAGAAAGAGAAATGTAATGTTTAAATTAACAAAGAAAGGATGATTAATGAT 1734

[illegible]

D	b		301	TTGAAAATTTAAGATATTAACTGAATAATGATGGAAAACTCAATATTAATTTACATTTA	360
O	y		3424	GAGAAGAAGAGCTCACACTACCTCTCAGCGGAGCATGCATCCACTGATCCGCCAGTCCAAG	3483
D	b		361	GGAAAGAAAAATTAATCTCTTCCTTATCAAGTGGAATTACATCATTTTAATCTGTAAATTA	420
O	y		3484	GAAGTCATTAAAGAACAGAACCTACCCGGCAATAGCCCAAGCAGATTAATACAGACTGG	3543
D	b		421	GAAGTAATTAATAAATTAATAATTAATACAGTAATTCCTCAAGTGAAATTAATAAGAAAGTT	480
O	y		3544	AATTAACGACACGGAATCTTACAGAAAGTCTCCCTGGAAGAACATATGCTCC-----	3598
D	b		481	AACGAAGCTTTAAATCTTACGAAAAATTTTTCTCCAGAAAGCAAAGTAATTCACAACAGTTGA	540
O	y		3597	-----CACTGGGTGTGTGAATCTGGCTCCGACACACTGGA	3633
D	b		541	ACTCCACTCAACCAAGATGTAAGTCTCCATCTCCATTATCTGTAAAGGTAAAGTGATGTTCA	600
O	y		3633	GCAGTCTCAACCTAAGA-----GCCATCTCACTCATGTGCGAGCCGAGTCCAAT	3684
D	b		601	GGATCCACAAAAGAAAGAACCAAAATACCAACTTAGCGCTTTATTAAACAGATTACAA	660
O	y		3685	ACAATTACCAATCTCAAGAACCTCGAGATAGAGTGGATGATCTACTTTGCTGCTATC	3744
D	b		661	CAGATGATCAATTAACAATAATTATGAGAAAGATGAATCTTACTTTATTAACCATTT	720
O	y		3745	TTCCGCGAGACCGAGGAGGATCAAGATGACCTCGGCGAGTGGTACCGGTGAGGCTGTC	3804
D	b		721	TTTGGAGATATCCGAAGATTAATGACGAATATTATGATCAAGTATTAATCTGGAAGCAATA	780
O	y		3805	ACTCCCTCCGATTTGATTAACATTTCTGTCMAAATCGAGAACGAATACGAATGCTCAT	3864
D	b		781	TCTGCACAATG---GATTAATCTCTCAGATTTGAAAAATGAATATGATGTATATAT	837
O	y		3865	CTGAACCTCTGCGAGCGCTATAGTCTCTCAAGAAACAGCTGGAGAAATAAGTGATG	3924
D	b		838	TTAAACCTTTAGCTGGAGATATAGAAAGCTTTAAAAACAAATTTGAAAAAAACATTTT	897
O	y		3925	ACCTCAATGTCAACGTGAAGAGCAATCTCAACAGCCGTTTAATATAGAGAGAAAAATTTC	3984
D	b		898	ACATTTAATTTTAATTTGAAACGATCTTAAATTCACGCTTATAGAAACGAAAAATATTTC	957
O	y		3985	AAGAAGCTTTGGAGACGACTTGATTCCTATTAAGACCTGCACTCTTAAGTACGTT	4044
D	b		958	TTAGATGATTAAGATCTGATTTATGCAATTTAAACATATATATCTCAATGAATACATT	1017
O	y		4045	GTCAGAGCCCATCAAGTCTCTCAATTAAGAAAGAAAGGATTAATTTCTGTCTAGTTC	4104
D	b		1018	ATTGAAAGTTCAATTTAATTAATTATGAAATTCGAAACAAAAAACACACTTTTAAAAAGTTAC	1077
O	y		4105	AACATATCAAGSACTCCATCGACACCGATATCAATTTGCTAATGATGCTGGGGAT	4164
D	b		1078	AAATATATAAAGATCAATGATGAATAATGATTAATTAATTTGCAAGAGAAAGTFAAGTTAT	1133
O	y		4165	TACAAGATCCAGACGGAATAATCAAGTCTGACCTTGACTATTAAAAAGTATTCAAC	4224
D	b		1138	TATGAAAGGTTTAGGGAATATTAAGAGATTTAGATTCATTTAAAAAATTTATCAAA	1197
O	y		4225	GATTAAG-----	4233
D	b		1198	GAAGAAAGAGATCCCATCATCAACCAACCAACACTCCGTCACACAGCAAAAAACAGAC	1257
O	y		4231	---CAAGCGGAGATGAAAAATATATCGCCCTTCCATTAATACATGCAAAACCTGTACAG	4287
D	b		1258	GAACAAAGAAAGGAAGTAAAGTCTTCCATTTTAAACAAACATTAAGACCTTAATACAT	1317
O	y		4288	ACAGTGAACGACAAATCGACCTTCTGTATTTACACTGGAGCCCAAGTCTCTCAACTAT	4347
D	b		1318	AACTAGTATTAATAATTTGACGATTAATTAATTAATTAACCTAAAGCAAAAGTATACGATTGT	1377
O	y		4348	ACTTACGAGAGAGCAATGTGGAGTTAAAAATCAAGAGCTGAATCTCAAAAACATC	4407
D	b		1378	AATGTTGAAAAAGATGAGAGCATGTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT	1437

Db 1018 ATTGAAGATTCATTAAATATATGTAATTCAGAACAAAAACACACTTTTAAAAAGTTAC 1077

Qy 4105 AACTATATCAAGAGCTCCATGACACCGATATCATTTCCGTAATGATGTGGGGTAT 4164

Db 1078 AAATATATTAAGAAATGATGTAATATATTAATTTGACACAGGAAGTATTAAGTAT 1137

Qy 4165 TACAGATCTGAGGAAAAATACAACTGACCTGACCTGATATTAATAAGATATCAAC 4224

Db 1138 TATGAAAAGTTTATGCAAAATATAGATGATTTAGATCATTAATAAAAGTTATCAA 1197

Qy 4225 GATAG----- 4230

Db 1198 GAGAAAAAGAAAGTTCCATCATCACACCAACACCTCCGTCACGCAAAAAA 1257

Qy 4231 -----CAAGCGAGATGAATAAATATGCTGCCCTCTGAAATTAATGAAACCGCTGAC 4284

Db 1258 GACGACAAAAAGAAAGTAAGTTCCCTTCATTTTAAACAAATGAGACCTTATAC 1317

Qy 4285 AAGACAGTGAACGACAAAAATGACCTCTGTAATTCACCTGAGCGCAAGCTCTCAAC 4344

Db 1318 AATACTTAATTAATAAATATGACGATTAATTAATTAAGCAAAAGATTAACGAT 1377

Qy 4345 TATACTTCGAGGAAGAGATGTAATTAATAATCAAGAGCTCACTACTCAAAACA 4404

Db 1378 TGTATATGTTGAAAAAGATGAAGCACATGTATAAATTAACCTAGATTTAAAGCA 1437

Qy 4405 ATCCAGACAGCTGCGAGATTTTCAGAAAAATTAACAATTCGCTGGAATTCAGACCGG 4464

Db 1438 ATGTATGCAAAATATGATCTTTTAAAAACCTTACGACTTCGAAGCAATTAATAAATG 1497

Qy 4465 TCTACCGATTAATACACACAAATCTCTGACCAAGTTTCTGTCACCTGAGCGATGCTTC 4524

Db 1498 ATAAATGATGATACGAAAAAAGATATGCTTGCAAAATTAAGTATAGTACAGGATTAAG--TT 1554

Qy 4525 GAAAAACCTCGGCAAAACAGCTGCTGAGCAATGCTCTGACGAGCAACCTGAGGCGATGCTG 4584

Db 1555 CAAAATTTTCTTAATACAAATATATCAAAATTAATTAAGGAAAAATTCCAAGATATGTTA 1614

Qy 4585 AACATCTCCAGCACCAATGCTGTAAGAAAGAGTCCGCCAGAAATAGCGGCTGTTTCAGG 4644

Db 1615 AACATTTACAAACCAATGCTGTAAGAAAGAGTCCGCCAGAAATTTGCGATGTTTCAGA 1674

Qy 4645 CATCTGACGAGCGGAGAGTGCATGCTCTGAACTCAAAACAAGAGAGATAG 4704

Db 1675 CATTTAGATGAAGAGAGATGTAATGTTTATTAATTAACAACAGAAAGTGATATAA 1734

Qy 4705 TGCGTGAAGAACCAACCAATGCTGCAATGAAACCAATGCGGCTGAGCGCCGATGCT 4764

Db 1735 TGTGTGAAAAATCCAAATCTACTTGTAAAGAAAAATATGATGATGACAGATCC 1794

Qy 4765 AAATGACAGGAGAGAGACAGCGGCTCTAAGGAAAAAATACATGCGAGTGTACTAG 4824

Db 1795 ACATGTACCGAAAAAGATTCAGGTAGCAGCAAGAAAGAAATACATGGAATGTACTAAA 1854

Qy 4825 CCGGACTCTCTACCTGCTGACGCGGATTTTTCCTCCAGCTCAATTTCTCGGCGATC 4884

Db 1855 CCGTATTTCTTATCCACTTTTGGATGATTTTTCGCACTTCTCTAACTTTCTTAGAATA 1914

Qy 4885 TCTTTCTGCTGATCTCATGCTGATCTCTGACAGCTTCATCTAA 4929

Db 1915 TCATTTCTTATTAATACATCATGTTATATATATACAGTTTCATTTAA 1959

RESULT 45

LOCUS AF061146 1116 bp DNA linear INV 09-FEB-2001

DEFINITION Plasmodium falciparum isolate HN6.24 merozoite surface protein 1

ACCESSION AF061146

VERSION AF061146.1 GI:3821984

KEYWORDS Plasmodium falciparum

SOURCE ORGANISM Plasmodium falciparum

REFERENCE 1 (bases 1 to 1116)

AUTHORS Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and Pluschke, G.

TITLE Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania

JOURNAL Acta Trop. 74 (1), 51-61 (2000)

MEDLINE 20106724

PUBMED 10643908

REFERENCE 2 (bases 1 to 1116)

AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.

TITLE Direct Submission

JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland

FEATURES

source

1..1116

/organism="Plasmodium falciparum"

/isolate="HN6.24"

/db_xref="taxon:5833"

/chromosome="9"

/note="wild isolate"

67..>1116

/gene="mspi"

67..>1116

/gene="mspi"

/product="merozoite surface protein 1"

67..>1116

/note="p190; surface protein with spectrin binding activity"

/codon_start=1

/product="merozoite surface protein 1"

/protein_id="AAC69745.1"

/db_xref="GI:3821985"

/translation="MKIIFFLCSELPFIINTQCVTHESYDELVKLEDAVLTVYS LFOKEWVLENGISGTAVTSTPGSGSVTSQSGSVASVAVSAGSGNS RRTNPSNDSBDAKSYADLKHRYQNTLTRELYPELFDLTNHLMLTDNIGFY LIDGEEINELLYLNFYUDELRAKLNDACANSYCOIPEFNKIRANLELDVYKIVFGY RKPLDNTKDNQKMEYIKRKTITIANINIELSKRTIQNKRAKNDDEGKKLYXO YNLFYIKQIQDEAHNLISVLEKRIDTLKKENIKRLDKIDEINPPANSNGTNPNTL LDKNKEEHEKEIKE"

BASE COUNT 458 a 136 c 161 g 361 t

ORIGIN

Query Match 10.7%; Score 528.2; DB:3; Length 1116;

Best Local Similarity 69.4%; Pred. No. 3.3e-105;

Matches 741; Conservative 0; Mismatches 308; Indels 18; Gaps 1;

Qy 10 ATGAATAATCATTTTCTCTCTGTCATCTTCTTTTATCATCACTACGTCG 69

Db 67 ATGAAGATCATATTTCTTTTATGTTTCATTTCTTTTATATATATACACATGTGTA 126

Qy 70 ACCACGAATCCTATCAGAGCTGTTAAGAAATCGAAGCTTTGGAAGATCCGTCCT 129

Db 127 ACACATGAAGATTAACAAGCTGTCAAAAAGCAAGAGCTTTGAAGAGCAAGTATG 186

Qy 130 ACCGATACAGCCTGTTCCAGAGGAGAAAGATGCTGTAATGAAGGACAGTGGCAG 189

Db 187 ACAGGTATATAGTTTATTTCAAAAGAAAAATGATTAATGAAGGAAACAAGTGGACA 246

Qy 190 GCCGTTACACAGCAGACCGCGCTCTAAGAGGTGTGGCTAGCGAGCGCGCGG 249

Db 247 GCTGTTACACTAGTACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306

Qy 250 TCTGTGCGCTGTGGGGGTTCCGTCGCTCCGCGGCGACGTCGGCATCAGTGGCTCAGTG 309

Db 307 TCAGGTGCTCAGGTGCTCAGT-----TGCCTCAGTTGCTCACTT 348

Qy 310 GCAAGCGCGGTTCCGGGAACAGTGAAGAACATCATCTGACACTATGACGATTC 369

Db 349 GCTTCAGGTGCTCAGGTCAAGTCAAGACGCTACAAATCCTTCAGATTAATCAAGTGAATCA 408

QY	370	GAGCGAAGTCTTACGGCGACTCGAAGCAGCGAGAGAACTATCTCTACTATCAAG	429
Db	409	GATGCTTAATATCTACGCTGATTTAAACGTAGAGTTCAAAATTTACTTGTCTACTATTA	468
QY	430	GAGCTGAAGTACCCACAGTGTGTTCGACCTCACTAATCATATCTGACACTGTGTGTAAC	489
Db	469	GAATCTCAAAATATCCGGAACCTTTTGATTTAACCAATCATATGTTAACTTGTGTGTAAT	528
QY	490	ATTCAATGCTTAAATATCTGATTCAGCGTTACGAAGAGATCAATAGACTCCTGTACAG	549
Db	529	ATTCAATGCTTCAAAATATTTAATGATGATATGAAGAAATTAATGAATATATATATAA	588
QY	550	TTGAATTTCTACTCTGCACTTGGTACGGGCAACTGAATGAGCTTGTGCGCAATGACTAT	609
Db	589	TTAAACTTTATATATGATTTATTTAGAGCCAAATTAATAGATGCAATGTCCATATGTTAT	648
QY	610	TGTCAAAATCCATTCATATTTGAAGATCAGAGCCAAACGAGCTTGACGTATTTGAAGAGTTG	669
Db	649	TGTCAAAATCCCTTTCAATCTTAAATTTCTGTGAAATGAATTTAGAGTACTTTAAAAAAT	708
QY	670	GCTTCGATTCGCAAGCCCTCGCAACATCAAGAGACATGTGGGAAGATGGAAGAT	729
Db	709	CGTTTCGGATATAGAAAACCATTAAGCATATTTAAAGATATATGTAGAGAAAATGGAAGAT	768
QY	730	TATATTTAAAGATATAGAGACCATCGAATCATTTACGAGCTGATGGAAGATCCAA	789
Db	769	TTCATTTAAAAAATTAATAAACACATAGCAAAATTAATTAATTTAATGAGAAAGTAAG	828
QY	790	AAGACCATAGCAAAAATATAGAAATGCAACGAGGAGGAGAAAGAAAGATGTACAG	849
Db	829	AAACCATATGATCAAAATATAGAAATGCAAGTATGAAAGAAAGGAAAAAATATATACAA	888
QY	850	GCCCAATACGACCTGTCCATATATACAACACAGCTTGAAGAAAGCCCATATACCTATCAGC	909
Db	889	GCTCATATATAACTTTTATTTATACATATTAACATTAACAAGACCATATTTTAAATAAAC	948
QY	910	GTAGCGAGAACCGGATAGACACCTCAAGAAAGATGAATATGAAATCAACTGCTCGAC	969
Db	949	GTTTATAGAAAAACGATTTGACACTTTAAAAAATAATGAATAATAGAAATTTACTTGAT	1008
QY	970	AAGATTAATGAATTAAGATCTCTCCGCGACCACTCTGGAGACACCCCTAACGCGT	1029
Db	1009	AAGATAGATGAATTAATAATATCCCCACGCGCAATTTCTGGAATATACCAATATCTCTC	1068
QY	1030	CTGGACAGAACAAAGAATAGAGACGACGAGAAAGAAAGATCAAGA	1076
Db	1069	CTTGATTAAGACAAAAAATCGAGAACGACGAAGAAAAATTAAGA	1115
RESULT	46		
PFHB3A			
LOCUS		1009 bp	DNA
DEFINITION		Plasmodium falciparum mezozoite surface antigen (MSA-1) gene	
LOCUS		fragment (clone HB3A).	
ACCESSION		X52962.1	GI:467603
VERSION		K52962.1	
KEYWORDS		mezozoite surface antigen; MSP1 gene; surface antigen.	
SOURCE		Plasmodium falciparum.	
ORGANISM		Plasmodium falciparum.	
REFERENCE		EduarVola; Alveolate; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS		1 (bases 1 to 1009)	
TITLE		Ranford-Cartwright,L.C.	
JOURNAL		Submitted (10-May-1990) Ranford-Cartwright L.C., Institute of Cell,	
REMARK		Animal and Population Genetics, University of Edinburgh, West Mains	
REFERENCE		Road, Edinburgh, EH9 3JN, Scotland	
AUTHORS		revised by [3] MAY	
TITLE		2 (bases 1 to 1009)	
JOURNAL		Ranford-Cartwright,L.C., Balfe,P., Carter,R. and Walliker,D.	
REMARK		Direct sequencing of enzymatically amplified DNA of alleles of the	
REMARK		mezozoite surface antigen MSA-1 gene from the malaria parasite	
JOURNAL		Plasmodium falciparum	
REMARK		Mol. Biochem. Parasitol. 46(1), 185-187 (1991)	

FEATURES	Source
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILALDELAVLYGSLFQEKMYLNEGSCAAVTT STPSKGSVASGGSGSVASGGSVSGSGSRRTNPDSNDSDASKYADLRKRVNR YLTLIKELKYPQLFDLTNHHMLTLDNTHGKRYLIDGEEINELLYKLNFFDLRLAL NDVCANDYCOIPENMLKIRANELDVLAKLVGKRYRPDLNIDNWKMDYIKKKNTIE NINLIEGSKRTIDKRNKNATPEEKRLYQAOYDLSTYNKOLEBAHNLISYLEKRIDPTD LKKNKIKLKLDDINEIKNPNSGNTPTLDDKNKKIEBEHEKEIKIAKTIFKFNID
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILALDELAVLYGSLFQEKMYLNEGSCAAVTT STPSKGSVASGGSGSVASGGSVSGSGSRRTNPDSNDSDASKYADLRKRVNR YLTLIKELKYPQLFDLTNHHMLTLDNTHGKRYLIDGEEINELLYKLNFFDLRLAL NDVCANDYCOIPENMLKIRANELDVLAKLVGKRYRPDLNIDNWKMDYIKKKNTIE NINLIEGSKRTIDKRNKNATPEEKRLYQAOYDLSTYNKOLEBAHNLISYLEKRIDPTD LKKNKIKLKLDDINEIKNPNSGNTPTLDDKNKKIEBEHEKEIKIAKTIFKFNID
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILALDELAVLYGSLFQEKMYLNEGSCAAVTT STPSKGSVASGGSGSVASGGSVSGSGSRRTNPDSNDSDASKYADLRKRVNR YLTLIKELKYPQLFDLTNHHMLTLDNTHGKRYLIDGEEINELLYKLNFFDLRLAL NDVCANDYCOIPENMLKIRANELDVLAKLVGKRYRPDLNIDNWKMDYIKKKNTIE NINLIEGSKRTIDKRNKNATPEEKRLYQAOYDLSTYNKOLEBAHNLISYLEKRIDPTD LKKNKIKLKLDDINEIKNPNSGNTPTLDDKNKKIEBEHEKEIKIAKTIFKFNID
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILALDELAVLYGSLFQEKMYLNEGSCAAVTT STPSKGSVASGGSGSVASGGSVSGSGSRRTNPDSNDSDASKYADLRKRVNR YLTLIKELKYPQLFDLTNHHMLTLDNTHGKRYLIDGEEINELLYKLNFFDLRLAL NDVCANDYCOIPENMLKIRANELDVLAKLVGKRYRPDLNIDNWKMDYIKKKNTIE NINLIEGSKRTIDKRNKNATPEEKRLYQAOYDLSTYNKOLEBAHNLISYLEKRIDPTD LKKNKIKLKLDDINEIKNPNSGNTPTLDDKNKKIEBEHEKEIKIAKTIFKFNID
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILALDELAVLYGSLFQEKMYLNEGSCAAVTT STPSKGSVASGGSGSVASGGSVSGSGSRRTNPDSNDSDASKYADLRKRVNR YLTLIKELKYPQLFDLTNHHMLTLDNTHGKRYLIDGEEINELLYKLNFFDLRLAL NDVCANDYCOIPENMLKIRANELDVLAKLVGKRYRPDLNIDNWKMDYIKKKNTIE NINLIEGSKRTIDKRNKNATPEEKRLYQAOYDLSTYNKOLEBAHNLISYLEKRIDPTD LKKNKIKLKLDDINEIKNPNSGNTPTLDDKNKKIEBEHEKEIKIAKTIFKFNID
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILALDELAVLYGSLFQEKMYLNEGSCAAVTT STPSKGSVASGGSGSVASGGSVSGSGSRRTNPDSNDSDASKYADLRKRVNR YLTLIKELKYPQLFDLTNHHMLTLDNTHGKRYLIDGEEINELLYKLNFFDLRLAL NDVCANDYCOIPENMLKIRANELDVLAKLVGKRYRPDLNIDNWKMDYIKKKNTIE NINLIEGSKRTIDKRNKNATPEEKRLYQAOYDLSTYNKOLEBAHNLISYLEKRIDPTD LKKNKIKLKLDDINEIKNPNSGNTPTLDDKNKKIEBEHEKEIKIAKTIFKFNID
REFERENCE	1. 1009
REFERENCE	/organism="Plasmodium falciparum"
REFERENCE	/db_xref="taxon:5833"
REFERENCE	/chromosome="9"
REFERENCE	/clone="HB3A"
REFERENCE	1. 1009
REFERENCE	/gene="MSP1"
REFERENCE	<1..>1009
REFERENCE	/gene="MSP1"
REFERENCE	/codon_start=1
REFERENCE	/product="merozoite surface protein 1"
REFERENCE	/protein_id="CA37135.1"
REFERENCE	/db_xref="GI:467604"
REFERENCE	/db_xref="SPRMBL:025865"
REFERENCE	/translation="MHSYQGLYKKILAL

D	b		265	GATGCTAAATCTTACCGTGATTTAAAACAAGACAGATACGAATAATTACTGTTAACATATCAAA	324
O	y		430	GAGCTGAAGTACCACAGTGTGGTGCACCTCACATATCATATCTGCACACATGTGTGATTAAC	489
D	b		325	GAACTCAAAATATCCCAACTCTTTCATTTAAGTATATCATATCATATGTMACTTGTGTCATAT	384
O	y		490	ATTCAATGCCCTCAATATATCTGATTTGACGGTTTACGAAGAAGATCAATGAACTCCTGTACAG	549
D	b		385	ATTCATGGTTTCCAATATATTTAATTCATGATGATATGAAATTAATTAATTAATTAATTAATTA	444
O	y		550	TTCGAATTCATCTTCGATGCTGCTAAGGGCCAAACTGATGATGACGTTTGGCCCAATGACAT	609
D	b		445	TTAAATCTTTATTTTGGATTTATTAAGACAAAAATTAATTCATGTATGTGCTATATGATTTAT	504
O	y		610	TCTCAATTCATTCATTTGATTTGAAGATTCAGAGCCAACGAGTTGGACGTATTTGAAGAAGTTG	669
D	b		505	TCTCAAAATACCTTTCAATCTTAAATTTGCTGCAAAATGAAATTAGACGTACTTAAAAAACTT	564
O	y		670	GTCCTCGGATATTCGACACCCCTCTGCACAAACATCAAGACATATGTGGGAAAGATGGAAT	729
D	b		565	GTTGTCGGATATGAAAAACCATTAACCAATTTAAAGATTAATGTAGGAAAAATGGAAGAT	624
O	y		730	TATATTTAAAGAAATPAAGACGACCTGCAAGAACATTAAGAGCGTATGCAAGAAATCCAAA	789
D	b		635	TTCATTTAAAAAAATPAAAAAACCTAGAAANTATTAATGATTTATTTGAAGGAAGTAAAG	684
O	y		790	AAGACCATAGACAAAAAATTAAGATTCGCAACCAAGAGAGAAAAAGAAAGTGTTCACAG	849
D	b		685	AAAACAATTTGATTAATAAATAATGATGTCACATTAAGAGAGAAAAAATAATTAACCAA	744
O	y		850	GCCCCAGTACGACCTGTCCATCTATTAACAACAACGCTTGAAGAGCCCCATPACCTCATCAGC	909
D	b		745	GCTCAATATGATCTTTCTATTATTAACAATTAACCAATTTAGAGAACGACATAAATTTAATAAGC	804
O	y		910	GTACGCGAAGCGCATAGACACCTCAAGAAATGAAATATCAAAAGAACTGCTGCAC	969
D	b		805	GTTTATGAAAAACGATTTAGACCTTTAAAAAATAATGAAAAATCAATAAATAATTAATTTAT	864
O	y		970	AAGATTATGAAATTAAGAAATCTCCGCGACCAACTTGGGAAACACCCCTCAACACGCTG	1029
D	b		865	AAGATTAATGAAATTAATAAATATCCGCCACGGCCAAATTTGGAAATATCACCAATATCTTC	924
O	y		1030	CTGCAACAAGAACAGAGATGAGGAGCAGACGAGAAAGAGATCAAGAGATGCGCCAAAAC	1089
D	b		925	CTTGATTAAGAACAAAAAATCGAGGACACGAAAAAGAAATTAAGAAATTTGCCAAAAC	984
O	y		1090	ATTAGTTCACATATGATTTCTCT	1114
D	b		985	ATTAAATTTAATTTGATGATTTAT	1009
RESULT 47					
AF061136					
LOCUS					1150 bp DNA linear INV 09-FEB-2001
DEFINITION					Plasmodium falciparum isolate IFa9.2 merozoite surface protein 1
ACCESSION					(msp1) gene, partial cds.
VERSION					AF061136
KEYWORDS					AF061136.1 GI:3821964
ORFANISM					Plasmodium falciparum.
SOURCE					Plasmodium falciparum
REFERENCE					Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
AUTHORS					Jiang,G., Daubenberger,C., Huber,W., Matile,H., Tanner,M. and Pluschke,G.
TITLE					Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
JOURNAL					Acta Trop. 74 (1), 51-61 (2000)
MEDLINE					20106724
PUBMED					10643908
REFERENCE					2 (bases 1 to 1150)

AUTHORS Jiang, G., Daubenberger, C.A., Matile, H. and Pluschke, G.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical
 Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES Location/Qualifiers
 source 1..1150
 /organism="Plasmodium falciparum"
 /isolate="IFA9.2"
 /db_xref="taxon:5833"
 /chromosome="9"
 /notes="wild isolate"
 71..>1150
 /gene="msp1"
 71..>1150
 /mrna
 /gene="msp1"
 71..>1150
 /product="merozoite surface protein 1"
 71..>1150
 /notes="p190; surface protein with spectrin binding
 activity"
 /codon_start=1
 /product="merozoite surface protein 1"
 /protein_id="AAC69735.1"
 /db_xref="GI:3821965"
 /translation="MKIIFPLCSFLPIINOCYTHESYOELVKKLEALDELAVLTGYG
 LFEHKNWLNEGTSIGTAVTTSTPSSKSSVTSGGSGSVASGSGSVASGSGSVASGSGS
 VASGSGNSRRTPNSDSDSAKSYADLKHRYONYLPTTIELKYPFLFDITNMHLL
 SNLHGFERYLIDGYEEINELLYKLNFEDLLRALINDVCANDYCOIFPNLKRANELD
 VLKLVFGYRKLPLDINKDNVGMEDYIKKKNTPTANINELIGSKRTIDONKNADNEE
 GKKLYOAYDLSYINKOLEAHNLISVLEKRIDLTKENIKLLEDIDIKITDAEKP
 TTESKPTLLDKNKIKIEHEKIKET"
 TSESPTLLDKNKIKIEHEKIKET"
BASE COUNT 467 a 139 c 176 g 368 t
ORIGIN
 Query Match 10.7%; Score 527.2; DB 3; Length 1150;
 Best Local Similarity 68.8%; Pred. No. 5.5e-105;
 Matches 743; Conservative 0; Mismatches 328; Indels 9; Gaps 1;
 Db 10 ATGAAATCATTTCTTCCTCTGTTTCATTTCTGTTTTATCATCATCAATCACTGAGTGC 69
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 71 ATGAGATCATATCTCTTTTATGTTCAATTTCTTTTATTTATTAATAACAAATGTGTA 130
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 70 ACCCAGCAATCCTATCAGAGAGCTGGTTAAGAACTGGAAGCTTTGGAAGATGCCGCTCT 129
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 131 ACACATGAAGATTATCAAGAGAGCTGTCTCAAAAAACTGAAAGCTTTAAGAAAGATCACTATTG 190
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Oy 130 ACCGATACAGCGCTTCCAGAGAGAAGATGGTGCAGTAAGGAAGGACGAGTGGCAGC 189
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 131 ACAGGTTATGGTTATTTTCATATAGAAAAAATGGATTAAATGAAAGAACAAATGTGACAA 250
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Oy 190 GCCGTTACAAACACACACACCCGGTCTTAAAGAGTCTGTGCGTAGCGGTCCGGTGGG 249
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 251 GCTGTTAACACATCAGTACACCTGTGTTCAAAAGGTTCAAGTTACTTTCAGGTGGTTCAGTGTGT 310
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 250 TCTGTGGCCTCTGGGGGT-----TCCGTCCGCTCCGGCGGACGCGTGGCATCAGGT 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 311 TCAGTTGCTTCAGGTGGTGTACAGTGGTCAAGTGGTTCAGTGGCTCAGTTGCTTCAAGT 370
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 301 GGCTAGTGGGCAAGCGCGGTTCGGGGAAACAGTGAAGAACCAATCCATCTGACAACTCT 360
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 371 GGCTCAGTTGGTTTCAGGTGGTTCAGGTAAATTAACAAGACGTCAAAATCTTCAGATTAATTTCA 430
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 361 AGCGATTCCGAGCGCAATCTCTACGCGCACACCTCAAGCACAGAGTAGAAGAACTATCTCTCT 420
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 431 AGTGAATTAGATGCTAAATCTTTACGCTGATTTAAACATPAGAGTTCAAAATTTACTTGTTC 490
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 421 ACTATCAGAGAGCTGAAGTACCCACAGTTGTTTCAGACCTCACTAATCATATCTGACACTG 480
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 491 ACTATTAAAGAACTCAATATCCGGAACCTCTTTGATTAAACCAATCATATCTTAACTTTG 550
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Oy 481 TGTGATAACATTCATGCGCTTCAATATCTGATTGACGGTTACGAAGAGATCAATGAACATC 540
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 551 AGTGAATTAATTCATGCTTCAATATTTAATGATTTGATGTGAAGAAATTAATTAATTA 610
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

OY	910	GTACGCGAGAGCCGATGACACCCCTCAGAAAGATGAAATATCAAAAGATGCTCGAC	969
Db	935	GTTTATGAAAAACGTATTGACGCTTTAAAAAAAATGAAACATTAAGAAATTACTTGGT	994
OY	970	AAGATTATGAAATTAACAATCTCCGGCAGCAACTCTGGGAAACACCCCTTAACGCGTG	1029
Db	995	AAGATTAATGAAATTTAAAAATCTCCCGCGGCCAATTCGGAAATATACCAATTACTCTC	1054
OY	1030	CTGGCACAAGACAGAGATAGGAGACGCGAATAAGATCAAGA	1076
Db	1055	CTTGATTAGAACCAAAAAATCGAGGAAACGGAAGAAAAATTTAAAAGA	1101

RESULT 50	AF061135	LOCUS	DEFINITION
AF061135	1157 bp	DNA	linear
	Plasmodium falciparum isolate IFN9.16 merozoite surface protein 1 (msp1) gene, partial cds.		INV 09-FEB-2001

ACCESSION	AF061135
VERSION	AF061135.1
KEYWORDS	GI:3821962
SOURCE	.
ORGANISM	Plasmodium falciparum. Plasmodium falciparum

REFERENCE
AUTHORS
Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
1 (pages 1 to 1157)
Jiang, G., Daubenberger, C., Huber, W., Matile, H., Tanner, M. and

TITLE	JOURNAL	YEAR	PAGE
Pluschke, G. Sequence diversity of the merozoite surface protein 1 of <i>Plasmodium falciparum</i> in clinical isolates from the Kilombero District, Tanzania	Acta Trop.	74 (1),	51-61 (2000)

MEDLINE
 PUBLISHED
 2010/07/24
 106433908
 2 (bases 1 to 1157)
 Jiang, G., Daubenberger, C.A., Matile, H. and Pluscnke, G.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical
 JOURNAL
 Institute, Socinstrasse 57, Basel 4002, Switzerland
 09502623

FEATURES	LOCATION/QUALITIES
SOURCE	1. .1157

```

/organism="Plasmodium falciparum"
/isolate="IFA9.16"
/db xref="taxon:5833"

```

```
/chromosome="9"  
/note="wild isolate"
```

```
gene      / 2. .>115/  
          /gene="mspl"  
mRNA     72. .>1157
```

```

/gene="msp1"
/product="merozoite surface protein 1"
72
>1157
cns

```

```

name="mspi"
/gene="mspi"
/note="p190; surface protein with spectrin binding

```

```
activity"
/codon_start=1
/product="merozoite surface protein 1"
```

```

/protein_id="AAC69734.1"
/db_xref="GI:3821963"
/defn="Protein encoded by the htr1a gene, which encodes a 5-HT1A receptor."

```

```
/classification= mliflucsfefrlnvcvfhstqylkvnlealedeavllgls  
lfhkemilneeitrtkgasogstfsgpsgsppssrsntlpnsntsga  
spdpadssdaakyaadlkhrvrnylllttkelkpqlfdltnmhtlclcdnhgfkylid
```

GYEEINELLYKLFYEDLLRAKLDVCAVDYCOJPENLKIRANELDVLKKEVGYRRE
LDNITKDVKMEDIKKNKKTIEENIEESKTTIDKRNKATKEEEKKLLYOQYDYL
LYNKOEEAHNLLSYLEKRIIDPTKKNKIKETLIDKINEJKNPPANGKNPTIIDDK

BASE COUNT	489 a	143 c	155 g	370 t
NKIIEHEKEIKIARTIKFNIDSLFTD"				

Query Match	Score	DB 3	Length
10.3%	508.2	1157	

Best Local Similarity 67.2%; Pred. No. 8.2e-101;
Matches 748; Conservative 0; Mismatches 338; Indels 27; Gaps 1;

10 ATGAAGAATCATTTTCTCCTCGTCGTCATTCCTGGTTTTATCAACAAATACAGGCGTG 69
OY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 ATGAGATCATCATCTCTTTATGTTCATCTCTTTTATTATAAATACACAATGTGTA 131
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 09 ACCAAGCAATCTCATCAGACCTGGTTAAAGAATTGGAAGCTTGGAAAAAGCCGCCIT 129
 | | | | |
L32 ACACATGAAGTTATCAAGAACTTGTCAAAACACTAGAACGTTAGAMATGCAGTATTG 191

130 ACCGGATACAGCCCTGTTCCAGGAAGGAGAAGATGCTGCTGATGAAAGGACGAGTGGCAG 189

Db 192 ACAGGTATAGTTTATTCATAAGGAAAAAATGATCTTAANTGAAGAAGAAATT----- 245

QY	190	GCCGTACAACCGACACACCCGGTCTTAAAGGCTCTGGCTACGCGGTGGCTCCGGTGGG	249
Db	246	-----ACTACAAAAGTGCACAGTGTCTCAAAGTGGTACAAAGTGT	284

250 TCTGAGCCCTGTGGGAGTCCGTCGCCCTCCGGGGCAGCGTGGCATCAGGTGGCTCAGTG 309

Db 285 ACAAGTGTACAAAGTGTCCAAAGTGTACAAAGTCCATCATCTGTTCAAC 344

310	GCAAGCGGCGGTCCGGGAAACAGTCGAAGAACCAATCATCTGTGCAACTGTAGCGATTCC	369
QY		
Db		
345	ACATTACCTCGTTCGAATACATTCATCTGTGGTGAAGCCCTCCACAGTGATGCAAGCATTTCA	404

370 GAGCCCAAGTCTACGCCGAGCTCAAGCACCGAGTGAAGAACTATCTCTCACTATCAAG 429

Db 405 GATGCTAATCTTACGCTGATTTAAAAACACAGAGTACGAATTACTTGTTAATATCAA 464

430 GAGTGAAGTACCACAGTGTGCTGACCTCACAATCATATGCTGACACTGTGTGATAAC 489

490 ATTGATGGCTCAAAATATCTGATTGAGGTTAGAGAGATCAATGAACCTGTACAAG 549

Db 525 ATTCATGGTTCAAAATTATTATGATGATATGAAGAAATTATGAATTATTATATAA 584

QY 550 TTGAATTCTACTTCGACTGCTAAGSGCCAACTGATGACGTTTGCGCCATGACAT 609

D5 585 TTAAACCTTTATTTTCGATTTTATTAACGCCAAATTAATATGATGATGCTAATGATTTAT 644

610 TGTCAATTCCATTTCATTTTGAGATCAGAGCCACGACTGGACGATTGAGAGAATTG 669

Db 645 TGTCAATACCTTTCATCTTAAATTCGTGCAATGATTTAGACGCTACTTAAAAACTT 704

QY 670 GTCCTCGGATATGCGAAGCCCTCTCGACACATCAAGGACAAATGTGGGAAGCATGGAAACAT 729

DB 730 TATATTAAAAAGAAATPAGAAGACCAATCGAGAACTTAACGAGCTGATCGAAGAATCCMAA 789

Db 765 TACATTAAAAAATAAAAAACCCATAGAAAAATATAATGAATTAAATTGAGGAAAGTAAG 824

QY 790 AAGACCATGACAAAATTAAGCATGCAACCAAGGAGGAAGAAAGAGAGTTGTACCAG 849

0Y	850	GCCCAAGTACGACCTGTCCATCTATAACCAACACGCTTGAAGAAGCCCATMCTCATCAGC	909
Db	825	AAACCAATGTATAAAAATATAGGAATGCCAATTAAGGAAGCAAGAAAAAATAATATACCA	884

Db 885 GCTCATATGATCTTTTATTTCACATTAACATTAGAGACACATATTTAATAAGC 944

910 GTACGAGAAAGCGCATAGACACCCCTCAGAAGAAATGAAATATTCAAAGAACTGCTGCAC 969

Db 945 GTTTTAGAANAACGTATTGCACCTTTAAAAAANAATGAANAACATTAAAGSAATTACTTGAT 1004

Db 1005 AAGATTAATGAATTTAAAAATCCCCCAGCGGCATTCGGAATTCACCCAAATACCTC 1064

QY 1030 CTGGACAGACAGAGATAGAGGACGACGAGAAAGATCAAGAGATCGCCAAACC 1089

Db 1065 CTTGATTAAGAACAAAAAATCGAGGAACACGAAAAAGAAATTAAGAATTCGCCAAACT 1124

||||| 1123 bp DNA linear INV 09-FEB-2001
Db 1125 ATTAATTATATGATGATTATTTCTGAT 1157

RESULT 51
AF061145
LOCUS Plasmodium falciparum isolate HN5 merozoite surface protein 1
DEFINITION (msp1) gene, partial cds.
ACCESSION AF061145.1 GI:3821982
VERSION AF061145.1
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 1123)
AUTHORS Jhang,G., Daubenberger,C., Huber,W., Matile,H., Tanner,M. and Pluschke,G.
TITLE Sequence diversity of the merozoite surface protein 1 of Plasmodium falciparum in clinical isolates from the Kilombero District, Tanzania
JOURNAL Acta Trop. 74 (1), 51-61 (2000)
MEDLINE 20106724
PUBMED 10643908
REFERENCE 2 (bases 1 to 1123)
AUTHORS Jhang,G., Daubenberger,C.A., Matile,H. and Pluschke,G.
TITLE Direct Submission
JOURNAL Submitted (22-APR-1998) Molecular Immunology, Swiss Tropical Institute, Socinstrasse 57, Basel 4002, Switzerland
FEATURES
source
1..1123
/organism="Plasmodium falciparum"
/isolate="HN5"
/db_xref="taxon:5833"
/chromosome="9"
/note="wild isolate"
71..>1123
/gene="msp1"
71..>1123
/gene="msp1"
/product="merozoite surface protein 1"
71..>1123
/note="p190; surface protein with spectrin binding activity"
/codon_start=1
/product="merozoite surface protein 1"
/protein_id="PAC69744.1"
/db_xref="GI:3821983"
/translation="MKIIFLCSFLPIINTQCVTHESYOELVKLEALDEAVLTGYS
LFOEKAVLNBSGTAVTSTPSKSGSGSGSVASVSGSVASVSGSGSGS
RRTNPSNNSDSQAKSYADLKHQVNTFLTKLYBELDUTLHMLTLDNIHGFYK
LIDGFEINELTKLNTFYDLRAKLNDACANSYCOIPEFNKLTANELDVKLTIVFG
RKPLDNIKDVNGKMEYIKKNKTTIANINELIESKRTIDQNKVADNEEGKKLYOQ
YNLEIYNKQIQEAHNLISVLEKRIIDLTKKNENIKKLEIDIKITDAEKPTGSKPMT
LLDNKKRIEHEEKIKE"
BASE COUNT 467 a 135 c 164 g 357 t
ORIGIN

Query Match 10.2% Score 502.4; DB 3; Length 1123;
Best Local Similarity 68.5% Pred. No. 1.5e-99;
Matches 733; Conservative 0; Mismatches 316; Indels 21; Gaps 2;

QY 10 ATGAAATCATTTCTCTCTGCTGATTCGTTTATATCAATCAATACAGTCGCTG 69
Db 71 ATGAAGATCAATCTCTTTATGTCATCTCTTTTATATTAATACACAAATGTATA 130

QY 70 ACCACGAATCTATACAGACTGTGTAAGAACTGAAGCTTTGGAGATGCCGCTT 129
Db 131 ACACATGAAATATACAGAACTGTCAAAAACTAGAACGTTTGAAGATGCACTATTG 190

QY 130 ACCGATACACCTGTTCCAGAGAGAAATGCTGAATGAAGGACAGTGGCAG 189
||||| 1123 bp DNA linear INV 09-FEB-2001

Db 191 ACAGTATATGATTATTCACAAAGCAAAAAATGATATTAATGAAGAACAGTGACAA 250
QY 190 GCCGTTACACACAGCACCCGGTTCTAAAGGTCTGCGCTAGCGGTCCGGTGG 249
Db 251 GCTGTACACATAGTACACCTGCTTCACAAAGGTTTCAGGTGCTTCCAGTGGC 310
QY 250 TCTGTGGCTCTCGGGGGTCCGTCGCGCTCCGGCGCACGTGGCATCGAGTGGCTCAGTG 309
Db 311 TCAGTTGCTTCAAGTTGCTTCAAGTGGC-----TCAGTTGCTTCAAGTT 352
QY 310 GCAAGCGCGGTTCCGGGACAGTGAAGAACCAATCCATCTGACACACTGATGATTC 369
Db 353 GCTTCAGGTGGTTCAGTAATTTCAAGACGTCAAAATCCCTCAGATTAATCAAGTATCA 412
QY 370 GACGCCAAGTCTTACGCCGACCTCAGACACCGAGTGAAGAACTATCTCTCACTCAAG 429
Db 413 GATGTAAATCTTACGCTGATTTAAACATAGAGTTCAAAATTAATCTTCACTATTTAA 472
QY 430 GAGCTGAAGTACCCACAGTGTGTCACCTCACTAATCATATGTCACACTGTGATTAAC 489
Db 473 GACCTCAATATCCGGAACCTTTTGATTTAACCAATCATATGTTACTTTGTGATATAT 532
QY 490 ATTCATGCTTCAAAATATCTGATGACGTTAGCAAGATCAATGAACCTCCTGTACAG 549
Db 533 ATTCATGCTTCAAAATATTTAATGATGATGATGATGAAGAAATTAATGAATTTATATA 592
QY 550 TTGAATTTCTACTTGACTTGTGATGAGGCCAACTGATGACGTTGGCCCAATACAT 609
Db 593 TTAACTTTTATGATTTATTAAGACCCAAATTAATGATGATGATGATGATGATGAT 652
QY 610 TGTCAATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 669
Db 653 TGTCAATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 712
QY 670 GTCTTGGATATCGCAGCCTCTGACAAATCAATCAATGATGATGATGATGATGATGAT 729
Db 713 GTCTTGGATATCGCAGCCTCTGACAAATCAATCAATGATGATGATGATGATGATGAT 772
QY 720 TATATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 789
Db 773 TATATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 832
QY 790 AAGACATAGACAAAAATAGAAATGCAACCAAGAGAGAAAGAAAGAAAGTGTACAG 849
Db 833 AAACATATGATCAAAATAGAAATGCAATGATGATGATGATGATGATGATGATGAT 892
QY 850 GCCCATGACGACCTGCTCATATACAAACAGCTTGAAGAAAGCCATACCTCATACG 909
Db 893 GCTCATATATATCTTTTATTTTACAAATTAACAAATTAACAAATTAATTAATTAAT 952
QY 910 GTACGAGAGAGCGCATGACACCCCTCAAGAAAGATGAAGATTAATCAAGAACTGCTGAC 969
Db 953 GTTTAGAAAACGATTTGACCTTTAAAAAATGAAGAAATTAATCAAGAACTGCTGAC 1012
QY 970 AAGATTAAGAAATTAAGA---ATCCCTCCGACCACTGTGGAGACACCCCTAACAG 1026
Db 1013 GATATAGATTAATTAAGAAAGATGCGCAAAAACCACTACTGGAAGTAAACCAATACT 1072
QY 1027 CTGCTGACAAAGACAGATAGAGACACGAGAAAGAGATTAAGA 1076
Db 1073 CTCTGTATAGTAAGAACAAAAATGAGAGACAGAAAGAAATTAAGA 1122

RESULT 52
AF061144
LOCUS Plasmodium falciparum isolate HN5 merozoite surface protein 1
DEFINITION (msp1) gene, partial cds.
ACCESSION AF061144.1 GI:3821980
VERSION AF061144.1
KEYWORDS Plasmodium falciparum.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.

OY	229	GCTAGCGGTGGCTCCGGGAGGCTCTGTGGCCCTCTGGGGGTTCCGTGCGCTCCGGGGGCGAGC	288
Db	121	ACTTCAGGATGGTCTCACTTACTTCACAGTGCTTCACTTACTTCACTTGCTTCACTTCA	180
OY	289	GTGCGCATCAGGTGGTGTGAGTGGCAAGCGCGGGTTCGGGAAACAGTCCAGAAACCAANTCA	348
Db	181	GTGTGCTTCAAGTGTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA	240
OY	349	TCTGCAACACTGTAGCGATTCCGAGCCCAAGTCTTACGCCGACCTTAAGCACCAGTGTAGA	408
Db	241	TCAGTATTAATTAAGGATTCAGATCTTAATCTTAACCTTACGTGATTTAAAACTAAGATTTCA	300
OY	409	AACTTCTCTCTCACAATCAAGAGAGCTGAAGTACCACAGTGTGTGACCTCTCAATCAT	468
Db	301	AATTAATCTGTTCACTTATTAAGAACTCAAAATATCCCGAATCTTGTATTAACCAATCAT	360
OY	469	ATGCTGACACTGTGTGATTAACATTCATGGCTTCAAAATATCTGATTTAGCGGTTCAGAAAG	528
Db	361	ATGTTAACTGTGTGTGATTAATATTCATGTTTCAAAATATTAATTAATGATGATATAGA	420
OY	529	ATCAATGAACCTCTGTACAAAGTTGAATTTCTTACTTCGACTTGTCTAAGGCCAACTGAAT	588
Db	421	ATTAAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	480
OY	589	GACGTTTCCGCCCAATGACATATTTGCAAAATTCATTCATTTGAAATATGACATGACCAACGAG	648
Db	481	CATGATATGTGCTTAATGATATTTGTCAAAATACCTTTCAATCTTTAAATTTCTGTCAAATGAA	540
OY	649	TTGACGATATTTGAGAAAGTTGTGCTTTCGGATATTCGCAAGCCCTCTGACACATCAAGGAC	708
Db	541	TTAGACGTAATTTAAAAAACTGTGTGCGATATTAAGAAACCATTAATTAATTAATTAAT	600
OY	709	AATGTGGGAAAGATTTGAAGATTTATTTAAAAAGATTAAGAAAGCAATGAGACATTTAAT	768
Db	601	AATGTAGGAAAAATGGAAGATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAAT	660
OY	769	GAGCTGATTCGAAGAATTCACAAAGACCATATGACAAATAATGAATGACCAACCAAGAGAA	828
Db	661	GAATTAATTTGAAGAAAGTAAAGAAACATTAATTAATTAATTAATTAATTAATTAATTAAT	720
OY	829	GAAAGGAAAGATGTGTACCGAGGCCAGTACGACCTGTCCATCTATTAACAAACAGCTTGA	888
Db	721	GGAATAAAAAAATTTATACCAAGCTCAATATTAATCTTTTATTAATTAATTAATTAATTAAT	780
OY	889	GAAAGCCATTAACCTCATAGCGTACTGGAGAAAGCCATTAAGACACCCCTCAAGAAGATGA	948
Db	781	GAAAGCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	840
OY	949	AATATCAAGAAGATCTGTGACAAAGATTAATGAATTAAGAAATTCCTCGGACAGCAACTCT	1008
Db	841	AATATCAAGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	900
OY	1009	GGAAGACACCCCTTAACACGCTGC--TGGCAAGAAAGCAAGAGATAGAGAGACGACGAGAA	1065
Db	901	GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	960
OY	1066	GAGATCAAGAGATGCGCAAAACCTTAAGTTCAACATAGAT	1107
Db	961	AAATTAAGAAAGATTTGCCAAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1002
RESULT 57			
PFAMSAIAK			
LOCUS			
DEFINITION	PFAMSAIAK	1002 bp	DNA linear
ACCESSION	PFAMSAIAK	1002 bp	DNA linear
VERSION	M77723.2	GI:21040184	
KEYWORDS	T-cell epitope; major	merozoite surface antigen; tripeptide repeat.	
SOURCE	Plasmodium falciparum		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
	1 (bases 1 to 1002)		

AUTHORS	Jongwattiwes,S., Tanabe,K.,Nekazawa,S., Yanaqi,T. and Kanbara,H.
TITLE	Sequence variation in the tripeptide repeats and T cell epitopes in P190 (MSA-1) of Plasmodium falciparum from field isolates
JOURNAL	Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)
DOI	92228018
PUBMED	1373473
REFERENCE	(bases 1 to 1002)
AUTHORS	Jongwattiwes,S.
TITLE	Direct Submission
JOURNAL	Submitted (21-MAY-2002) Faculty of Medicine, Chulalongkorn University, Rama4, Pathumwan, Bangkok 10330, Thailand On May 21, 2002 this sequence version replaced gi:309712. Location/Qualifiers
COMMENT	1..1002
FEATURES	/organism="Plasmodium falciparum"
SOURCE	/db_xref="taxon:5833"
	/map="g"
	/clone="835B"
	/sex="asexual"
gene	/dev_stage="erythrocytic"
	<1..>1002
	/gene="P190"
CDS	<1..>1002
	/gene="P190"
	/codon_start=1
	/product="major merozoite surface antigen"
	/protein_id="AA029670_2"
	/translation="ALSDAVLTGYSLEFQEKVNLNSTGVTTSTPPSGSVTSGGG SVTSGGSVTSAVASVAASVASVGSSGNRRNPSPDNSDSDAKSYADLKHVVONY LEFTKEIKPELPFLDTNHMLTCINIGFKLFYGEIINELTKLNFPDLRAKLIN DYVNCDCOILPENLKIRANIELDIVAKLVFGVRKRLDNIDKNVGKMEKY IKKKKTIAN INELIESSKRTIIDNKNADEEGEKKRILOAOYNFTYINKOLOEAHNLISVEKRITDLL KKNNIKRLLDKIDEIKRPAPNGNTPNPLPKKKREVBHEKIKEIAKTIFENID "
BASE COUNT	414 a 130 c 155 g 303 t
ORIGIN	
Query Match	10.0% Score 496.4 ; DB 3 : Length 1002;
Best Local Similarity	69.2% Pred. No.3.2e-98;
Matches 693; Conservative 0; Mismatches 306; Indels 3; Gaps 1;	
Oy	109 GCTTGGAAGATGCCCGCTCTACCCGATACAGCCTGTTCAGAGAAGAGCATGTGCCTG 168
Dd	1 GCATTAGAAGATGCAGTATTGACAGCTTAATACTTATTTCAAAAGAAAATGTTATTA 60
Oy	169 AATGAAGGAGCAGTAGTGACAGGCGCGTTACAACCACACACCOCGTTCTAAAGGCTCTGTG 228
Dd	61 AATGAAGAACACAGTGAAGAAACAGCTTTACACTAGTACACCTGTTACAGTGTTCAAGTT 120
Oy	229 GCTAGCGGTGGCTCCGGTGGGTCTGTGGCCCTCTGGGGGTTCCGTGCTCCGGGGCACG 288
Dd	121 ACTTCAGGTGGTTCAGTACTTCAGGTGGTTTCAGTTACTAGTTGCTTCAGTTGCTTCA 180
Oy	289 GTMGCATGAGGTGGCTCAGTGSCAGGCGGGTTCGGGAACAGTCCGAAGAACCAATCCA 348
Dd	181 GTTGCTTAGTGGTTCAGTTGCTTCACAGGTGGTTCAGTAAATTCAGACGTTCAATCTCT 240
Oy	349 TCTGCACACTAGAGATTCCGACGCCCAAGTCTCAACCGCAGCACTCAACACAGGATGAGA 408
Dd	241 TCAGATAATTAAAGATATTCAGATCTTAATCTTACGCTGATTTAAACATRAGAGTTCAA 300
Oy	409 AACATATCTCTACATATCCAGAGAGCTGAAGTACCCACAGTGTTCGACCTCAATATCAT 468
Dd	301 AAATTACTGTTCACATATTAAAGAACTCAAAATATCCGAAACCTCTTGATTTAACCAATCAT 360
Oy	469 ATGCTGACACTGTGTGATTAACATTATGCGTTCAAATATCTGATTGACGGTTACGAAGA 528
Dd	361 ATGTAACTTTGTGTGATTAATATTCATGTTTTCAAAATTTTAAATTGATGATCAAGAA 420
Oy	529 ATCATATGAACTCGTACAAAGTTGAATTTTACTCTCGACTGCTGCTAAGGCGCAACATGAT 588
Dd	421 ATTAAATGAATTAATTAATAAATTAACCTTTATTTTGAATTTATTAAGACAAAAATTAAT 480

OY	589	GACGTTTGCGCCAAAGCATATTCGCATAATTCATCATATTGGAAATCAGGCCAACGAC	648
Dd	481	GATGTATGTGGCTAAAGATGATTATTGTCAAAATACCCTTTCGAATCTTAAAATTCGTGCCAATGAA	540
OY	649	TTCGACGTAATTCGAAACGTTGGCTCTTCGGATATTCGCCAAGCCCTCTGCACAACATCAAGAC	708
Dd	541	TTAGACGTACTTAAAAAACTGTGTTCGGATTTACAAAAACCTTTGACACTATTTAAAGAT	600
OY	709	AATGTGGCAAGATGGAAGATTATATTAATAAAGAATTAAGAACCATCGAGACATTAAAC	768
Dd	601	AATGTAGGAAAAATGGAAGATTACATTAAAAAAAATTAACACCATAGCAATATTAAT	660
OY	769	GAGCGTAGTCGAAGATTCACAAAAGACCATTAGACAAATAATAGAAATGCCAACGAGNAGAA	828
Dd	661	GAATTAATTTGAAGGAAGTAGAAGAAACAATGTGCATAAATAAGAAATGACAGATTAATGAGAA	720
OY	829	GAATAAGAAAGATGTTTACACGAGCCAGTAGACACTGTCCATCTATTAACAAACAGCTTGAA	888
Dd	721	GGAAAAAATAATTTTACCAAGCTCAATATATATCTTTTATTTACAAATPAACAATTTACAA	780
OY	889	GAGCCCCATACCTTCATCAGCGTACTGGAGAAAGCCATAGACACCCCTCAGAAGATGAA	948
Dd	781	GAGACGACATTAATTTATTAAGCGCTTTTAAAAAAACGTATTGACCTTTAAAAAAATGAA	840
OY	949	AATATCAAAGAACTCTGTGACAGATTAATGAATTAAGAAATCTCCGCCAGCCCAACTTC	1008
Dd	841	NACATPAAGAAATTTACTGTGATAGATTAATGAATTAATAAATCCCAACCGGCAATTCCT	900
OY	1009	GGGAACACCCCCTAACACCCCTGC---TGGACAAGAACAAGAAAGATAGAGAGCACGAGAAA	1065
Dd	901	GGAAATPACACCAATTCCTCTCTCCCTCAGAAATAGAAAAAAGAAAGTGGAGGACACGAGAA	960
OY	1066	GAGATCAAGAGATGCGCAAAAACCATTAAGTTCAACATGAT	1107
Dd	961	AAAAATMAAGAAATTCGCAAAAACATTTAATTTAATTAACATTGAT	1002
RESULT 58			
PFMSA1AB			
LOCUS		PFMSA1AB	1002 bp DNA linear INV 22-MAY-2002
DEFINITION		Plasmodium falciparum clone 807 major merozoite surface antigen (p190) gene, partial cds.	
ACCESSION		M77714.2	GI:21040166
VERSION		M77714	
KEYWORDS		T-cell epitope; major merozoite surface antigen; tripeptide repeat.	
SOURCE		Plasmodium falciparum	
ORGANISM		Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
REFERENCE		1 (bases 1 to 1002)	
AUTHORS		Jongkuntives S., Tanabe K., Nakazawa S., Yanaagi T. and Kanbara H.	
TITLE		Sequence variation in the tripeptide repeats and T cell epitopes in p190 (MSA-1) of Plasmodium falciparum from field isolates	
JOURNAL		Mol. Biochem. Parasitol. 51 (1), 81-89 (1992)	
MEDLINE		92228018	
PUBMED		1373473	
REFERENCE		2 (bases 1 to 1002)	
AUTHORS		Jongkuntives S.	
TITLE		Direct Submission	
JOURNAL		Submitted (21-May-2002) Faculty of Medicine, Chulalongkorn University, Rama4, Pathumwan, Bangkok 10330, Thailand	
COMMENT		On May 21, 2002 this sequence version replaced gi:309696.	
FEATURES		location/Qualifiers	
source		1..1002	
		/organism="Plasmodium falciparum"	
		/db_xref="taxon:5833"	
		/map="g"	
		/clone="807"	
		/sex="asexual"	
		/dev_stage="erythrocytic"	
		<1..>1002	
		/gene="p190"	
		<1..>1002	
CDS			

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 21:58:31 ; Search time 761 Seconds
(without alignments)
14618.754 Million cell updates/sec

Title: US-09-269-874a-2

Perfect score: 4940

Sequence: 1 cgcacgcgtatgaataatcat.....ttcatctaatagatcgtg 4940

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 180 summaries

Database :

N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001a.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001b.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	4938.4	100.0	4940	19	AAV21451	P. falciparum mod1
2	2837.6	57.4	4940	19	AAV35363	P. falciparum gp19
3	2706.2	54.8	5760	6	AAV50530	Sequence encoding
4	1262	25.5	5181	16	AAO80911	Plasmodium falcipa
5	1000	20.2	1950	18	AAAT97956	Chimeric MSA-1 ant
6	989	20.0	1897	18	AAAT97958	Chimeric MSA-1 ant
7	969.6	19.6	1896	18	AAAT97957	Chimeric MSA-1 ant
8	744.2	15.1	1065	20	AAK56008	Merzoite surface
9	744.2	15.1	1065	20	AAK25586	Merzoite surface

10	742	15.0	1140	20	AAK56021	Merzoite surface
11	742	15.0	1140	20	AAK25593	Merzoite surface
12	719.8	14.6	1077	22	AAK68976	Merzoite surface
13	625.2	12.7	1088	20	AAK56009	Merzoite surface
14	625.2	12.7	1088	20	AAK25587	Merzoite surface
15	525.2	10.6	786	22	AAK68978	Merzoite surface
16	322.8	6.5	1128	22	AAK68980	DNA encoding major
17	308.8	6.3	660	9	AAK81148	DNA sequence encod
18	308.8	5.0	1068	9	AAK81151	DNA sequence encod
19	246.6	5.0	354	18	AAK80403	PFMSPI (p19)A codin
20	246.6	5.0	354	18	AAK80403	PFMSPI (p19)A codin
21	243.4	4.9	737	6	AAK50355	Storage-specific,
22	240.2	4.9	668	9	AAK82176	31-1 Repeated Dele
23	201.8	4.1	456	9	AAK81150	DNA sequence encod
24	201.8	4.1	462	9	AAK81149	DNA sequence encod
25	198.4	4.0	333	22	AAK68977	Merzoite surface
26	194.8	3.9	387	18	AAK80404	PFMSPI (p19)S codin
27	194.8	3.9	387	18	AAK80404	PFMSPI (p19)S codin
28	194.6	3.9	309	24	AAK22459	Plasmodium falcipa
29	194.6	3.9	3147	24	AAK22460	Plasmodium falcipa
30	194.6	3.9	3147	24	AAK22461	Plasmodium falcipa
31	191.4	3.9	309	24	AAK22462	Human C3d3 DNA-PFM
32	191.4	3.9	3147	24	AAK22463	Mutant PFMSPI.19-h
33	191.4	3.9	3147	24	AAK22464	Storage-specific,
34	119.2	2.4	306	6	AAK50354	Chicken leucocytos
35	105.6	2.1	3399	17	AAK05868	DNA encoding signa
36	104	2.0	618	22	AAK05868	DNA encoding signa
37	99.4	2.0	165	18	AAK93729	Drosophila melanog
38	99.4	2.0	165	18	AAK93729	Drosophila melanog
39	85	1.7	5688	23	ABL26543	DNA encoding Leuco
40	85	1.7	5688	23	ABL26543	Synthetic gene VIV
41	81.6	1.7	7588	16	AAK07587	Synthetic gene VIV
42	81.6	1.7	7588	16	AAK07587	Synthetic gene VIV
43	72.4	1.5	5438	22	AAK47054	Plasmodium falcipa
44	72.4	1.5	5438	22	AAK47054	Plasmodium falcipa
45	70.6	1.4	3579	21	AAK47059	Oligonucleotide fo
46	70.2	1.4	646	24	ABO46953	Oligonucleotide fo
47	70.2	1.4	646	24	ABO46953	Oligonucleotide fo
48	68.6	1.4	969	24	ABO39480	Oligonucleotide fo
49	68.6	1.4	969	24	ABO39480	Synthetic gene VIV
50	67.2	1.4	5643	22	AAK47056	Synthetic gene VIV
51	65.2	1.3	8045	21	AAK65171	Neutrospora crassa
52	63.8	1.3	14006	24	ABL33958	Human immune syste
53	63.8	1.3	5163	19	AAK20700	Cryptosporidium pa
54	60.6	1.2	5163	21	AAK61849	Oligonucleotide fo
55	60.6	1.2	5163	21	AAK61849	Oligonucleotide fo
56	60.6	1.2	5318	21	AAK61848	Oligonucleotide fo
57	60.6	1.2	5318	21	AAK61848	DNA encoding a por
58	60.2	1.2	6668	24	ABL36377	C parvum GP900 gen
59	59.6	1.2	489	22	AAK00654	Human immune syste
60	59.2	1.2	557	24	ABO43458	Plasmodium vivax m
61	59.2	1.2	557	24	ABO43458	Oligonucleotide fo
62	59.2	1.2	557	24	ABO43458	Oligonucleotide fo
63	58.8	1.2	5318	19	AAK20701	Cryptosporidium pa
64	58.8	1.2	5318	19	AAK20701	Cryptosporidium pa
65	58.8	1.2	5511	21	AAK61847	C parvum GP900 gen
66	58.8	1.2	7334	21	AAK61846	Cryptosporidium pa
67	58.4	1.2	7334	24	ABK04775	C parvum GP900 gen
68	58.4	1.2	5535	21	AAK70184	Plasmodium falcipa
69	58.2	1.2	1380	21	AAK234650	Human growth facto
70	57.8	1.2	3252	24	AAK22448	Human immune syste
71	57.8	1.2	3252	24	AAK22448	Human immune syste
72	57.2	1.2	13631	23	ABL20334	B. burgdorferi ant
73	56.4	1.1	379	23	AAK590909	DNA encoding novel
74	56.4	1.1	379	23	AAK590909	DNA encoding novel
75	55.8	1.1	1998	21	AAK70212	Plasmodium falcipa
76	55.4	1.1	6522	24	ABL33050	Human immune syste
77	55.2	1.1	1039	20	AAK61812	B. burgdorferi ant
78	55.2	1.1	1125	20	AAK61811	B. burgdorferi ant
79	54.6	1.1	35515	20	AAK20252	Borrelia burgdorfe
80	54.6	1.1	2427	23	AAK69554	DNA encoding novel
81	54.2	1.1	2427	23	AAK68862	DNA encoding novel
82	54.2	1.1	9209	24	ABL34427	Human immune syste
			9219	22	AAK46808	Tumour suppressor

C	83	54	1.1	3662	23	ABL05274
C	84	53.6	1.1	2943	21	AAAT0229
C	85	53.2	1.1	1167	23	AAAT07022
C	86	53.2	1.1	1413	23	AAAT1483
C	87	53.2	1.1	1413	23	AAAT1483
C	88	53.2	1.1	1413	23	AAAT1483
C	89	53.2	1.1	1606	23	AAAT1483
C	90	53.2	1.1	1822	23	AAAT1483
C	91	52.8	1.1	6644	20	AAAT1483
C	92	52.8	1.1	6767	22	AAAT1483
C	93	52.8	1.1	7372	20	AAAT1483
C	94	52.8	1.1	7797	20	AAAT1483
C	95	52.8	1.1	7996	20	AAAT1483
C	96	52.4	1.1	752	24	ABO14854
C	97	52.4	1.1	752	24	ABO14854
C	98	52.4	1.1	1080	24	ABO14854
C	99	52.4	1.1	4593	24	ABO14854
C	100	52.2	1.1	1140	21	AAAT0071
C	101	52.2	1.1	1662	23	ABO14854
C	102	52	1.1	393	23	AAAT0684
C	103	52	1.1	9539	22	AAAT0684
C	104	52	1.1	9539	22	AAAT0684
C	105	51.6	1.0	863	22	AAAT0684
C	106	51.4	1.0	4590	22	AAAT0684
C	107	51.2	1.0	453	23	AAAT0684
C	108	51.2	1.0	484	24	ABO31102
C	109	51.2	1.0	647	24	ABO31102
C	110	51.2	1.0	647	24	ABO31102
C	111	51.2	1.0	5454	21	AAAT0236
C	112	51	1.0	822	24	ABO44852
C	113	51	1.0	822	24	ABO44852
C	114	50.8	1.0	621	23	AAAT0688
C	115	50.8	1.0	10732	21	AAAT0594
C	116	50.6	1.0	8392	24	ABO33490
C	117	50.2	1.0	556	24	ABO36996
C	118	50.2	1.0	556	24	ABO36996
C	119	50.2	1.0	4000	18	AAAT91902
C	120	50	1.0	5661	19	AAAT21518
C	121	49.8	1.0	1563	24	ABO33204
C	122	49.8	1.0	2336	23	AAAT5652
C	123	49.6	1.0	441	23	AAAT5652
C	124	49.6	1.0	887	22	AAAT94064
C	125	49.6	1.0	1330	16	AAAT92525
C	126	49.6	1.0	7503	21	AAAT0206
C	127	49.6	1.0	49999	20	AAAT23895
C	128	49.6	1.0	50000	21	AAAT64139
C	129	49.4	1.0	693	24	ABO32104
C	130	49.4	1.0	693	24	ABO32104
C	131	49.4	1.0	867	20	AAAT91460
C	132	49.4	1.0	867	20	AAAT91461
C	133	49.4	1.0	867	22	AAAT42783
C	134	49.4	1.0	867	22	AAAT42784
C	135	49.4	1.0	1200	24	ABO39210
C	136	49.4	1.0	1200	24	ABO39211
C	137	49.4	1.0	1397	20	AAAT91462
C	138	49.4	1.0	1397	20	AAAT91463
C	139	49.4	1.0	1397	22	AAAT42785
C	140	49.4	1.0	1397	22	AAAT42786
C	141	49.4	1.0	7442	22	AAAT4686
C	142	49.4	1.0	8079	24	ABO45148
C	143	49.2	1.0	541	24	ABO45149
C	144	49.2	1.0	541	24	ABO45149
C	145	49.2	1.0	807	22	AAAT21312
C	146	49.2	1.0	875	22	AAAT95044
C	147	49.2	1.0	29392	19	AAAT15422
C	148	48.8	1.0	571	23	AAAT90731
C	149	48.8	1.0	506	23	AAAT86116
C	150	48.8	1.0	906	23	AAAT86069
C	151	48.8	1.0	993	23	AAAT86118
C	152	48.8	1.0	993	23	AAAT86067
C	153	48.8	1.0	993	23	AAAT90706
C	154	48.8	1.0	1149	21	AAAT51336
C	155	48.8	1.0	2579	23	ABO26836

Drosophila melanog
Plasmodium falcipar
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Base sequence of t
Tumour suppressor
Base sequence of t
Compox virus bsr t
Base sequence of t
Oligonucleotide fo
Oligonucleotide fo
Mouse ischaemic co
Human liver glucoc
Arabidopsis thalia
Drosophila melanog
DNA encoding novel
Chemically pretrea
DNA transcription
Human neuroblastom
Yeast AOD9604-asso
DNA encoding novel
Consensus sequence
Oligonucleotide fo
Plasmodium falcipa
Oligonucleotide fo
Oligonucleotide fo
Gene encoding a su
Human immune syste
Oligonucleotide fo
Mannose-1-phosphat
Rattus norvegicus
Nucleotide sequenc
Drosophila melanog
DNA encoding novel
Human neuroblastom
Nicotiana glumabagi
Plasmodium falcipa
Murine LOBO homolo
Nucleotide sequenc
Oligonucleotide fo
Oligonucleotide fo
T. gondii immunoge
T. gondii nte867 r
T. gondii cdna enc
Reverse complement
Oligonucleotide fo
Oligonucleotide fo
T. gondii immunoge
T. gondii nte1397
Reverse complement
Tumour suppressor
Chemically treated
Oligonucleotide fo
Oligonucleotide fo
Human breast cance
Human neuroblastom
Mouse poly Ig rece
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Recombinant proteoi
Drosophila melanog

C	156	48.8	1.0	3231	23	ABL26858
C	157	48.8	1.0	8895	22	ABA06801
C	158	48.8	1.0	8895	22	ABA06801
C	159	48.8	1.0	8895	22	AAAT28888
C	160	48.8	1.0	8895	22	AAAT28888
C	161	48.8	1.0	9656	22	ABA06800
C	162	48.8	1.0	9656	22	ABA06800
C	163	48.8	1.0	9656	22	AAAT06847
C	164	48.8	1.0	9656	22	AAAT06847
C	165	48.8	1.0	14041	22	AAAT06847
C	166	48.8	1.0	16842	22	AAAT06847
C	167	48.8	1.0	16842	24	AAAT06847
C	168	48.8	1.0	16842	24	AAAT06847
C	169	48.8	1.0	16842	24	AAAT06847
C	170	48.8	1.0	16842	24	AAAT06847
C	171	48.4	1.0	792	18	AAAT25071
C	172	48.4	1.0	1612	7	AAAT60392
C	173	48.4	1.0	21354	22	AAAT6815
C	174	48.2	1.0	1690	16	AAAT92524
C	175	48.2	1.0	7536	24	ABO23312
C	176	48	1.0	727	24	ABO23312
C	177	48	1.0	727	24	ABO23312
C	178	48	1.0	2827	23	ABO23303
C	179	48	1.0	2827	23	ABO23303
C	180	47.8	1.0	390	23	AAAT59541
C	181	47.8	1.0	1684	23	AAAT90724

ALIGNMENTS

RESULT 1
ID AAV21451 standard; DNA; 4940 BP.
AC AAV21451;
XX 23-SEP-1998 (first entry)
DT P. falciparum modified gp190 DNA.
XX
KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW monoclonal antibody; passive immunisation; parasite; ss.
OS Plasmodium falciparum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 10..4929
FT /tag= a
FT /product= gp190
MO9814583-A2.
PD 09-APR-1998.
PF 02-OCT-1997; 97WO-BP05441.
PR 02-OCT-1996; 96DE-4040817.
PA (BUTJ/) BUTJARD H.
XX
XX Bujard H, Pan W, Tolle R;
XX
XX MPI: 1998-240088/21.
XX P-PSDB; AAV54145.
XX
XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
XX protein - useful in anti-malaria vaccines, also stabilising genes by
XX reducing their AT content
XX
XX Example 1; Fig 3c; 48pp; German.
XX
XX This sequence encodes a modified Plasmodium falciparum gp190/MSP-1

CC (merozoite surface protein) which has a reduced AT content resulting in
CC a higher stability of the protein. Such a protein is useful in
CC vaccines against malaria or for producing monoclonal antibodies (for
CC passive immunisation). The complete gp190 protein can now be produced
CC outside the parasite and has, at least over extended regions, the native
CC pattern of folding. Larger amounts of the protein can be produced
CC recombinantly than would be possible using the parasites as source.
XX

Sequence 4940 BP: 1669 A; 1156 C; 1060 G; 1055 T; 0 other;

Query Match 100.0%; Score 4938.4; DB 19; Length 4940;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4939; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Oy 1 CGCAGCGCTATGAAAAATCTCTCTCTGCTCATTTCTGTTTATCATCACTACT 60
Db 1 CGCAGCGCTATGAAAAATCTCTCTCTGCTCATTTCTGTTTATCATCACTACT 60
Oy 61 CAGTCCGTGACCCAGCAATCCTATCAGAGCTGTTAAGAACTGGAAGCTTTGGAAGT 120
Db 61 CAGTCCGTGACCCAGCAATCCTATCAGAGCTGTTAAGAACTGGAAGCTTTGGAAGT 120
Oy 121 GCCGTCTTACCGGATACAGCTGTTCAGAAAGAAAGATGTCTGAATGAAAGGACG 180
Db 121 GCCGTCTTACCGGATACAGCTGTTCAGAAAGAAAGATGTCTGAATGAAAGGACG 180
Oy 181 AGTGGCAGGGGCGTTACAAACGACACCGGTTCTAAAGGCTGTGGCTAGCGGTGGC 240
Db 181 AGTGGCAGGGGCGTTACAAACGACACCGGTTCTAAAGGCTGTGGCTAGCGGTGGC 240
Oy 241 TCCGGTGGGTCTGTGCGCTCTGCGGGGTTCCGTGCGCTCCGGGCGAGCGTGCATCAGGT 300
Db 241 TCCGGTGGGTCTGTGCGCTCTGCGGGGTTCCGTGCGCTCCGGGCGAGCGTGCATCAGGT 300
Oy 301 GGCCTACGTGGCAAGGGGGGTTCCGGGAAACAGTTCGAACCAATCCATCTGACACACT 360
Db 301 GGCCTACGTGGCAAGGGGGGTTCCGGGAAACAGTTCGAACCAATCCATCTGACACACT 360
Oy 361 AGCGATTCGAGCGCAATCTCTACGCGACCTCAAGCAGCAGAGTGAAGAACTATCTCTC 420
Db 361 AGCGATTCGAGCGCAATCTCTACGCGACCTCAAGCAGCAGAGTGAAGAACTATCTCTC 420
Oy 421 ACTATCAAGAGCTGAAATGACCCAGTGTTCGACCTCACTAATCATATGCTGACACTG 480
Db 421 ACTATCAAGAGCTGAAATGACCCAGTGTTCGACCTCACTAATCATATGCTGACACTG 480
Oy 481 TGTGTAACATTCATGCTTCAATATCTGATTCAGCGTTACGAAGATCAATGACTC 540
Db 481 TGTGTAACATTCATGCTTCAATATCTGATTCAGCGTTACGAAGATCAATGACTC 540
Oy 541 CTGTACAGTTGAATTTCTACTGACTGCTAAGGGGCAAACTGAAGAGCTTTGGCGC 600
Db 541 CTGTACAGTTGAATTTCTACTGACTGCTAAGGGGCAAACTGAAGAGCTTTGGCGC 600
Oy 601 AATGACTATTTGCAAAATCCATTCATTTGAAGATCAGAGCAAGAGTTGGACTATTTG 660
Db 601 AATGACTATTTGCAAAATCCATTCATTTGAAGATCAGAGCAAGAGTTGGACTATTTG 660
Oy 661 AAGAAAGTGTCTTCGATATCGCAAGCCTCTCGAACATCAAGAGCAATGTGGAAAG 720
Db 661 AAGAAAGTGTCTTCGATATCGCAAGCCTCTCGAACATCAAGAGCAATGTGGAAAG 720
Oy 721 ATGGAAGTTATTTAATAAAGATTAAGAACCATCGAAGCATTAAGAGCTGTATCAA 780
Db 721 ATGGAAGTTATTTAATAAAGATTAAGAACCATCGAAGCATTAAGAGCTGTATCAA 780
Oy 781 GAATTCAAAAAGACCATAGACAAAAATTAAGATGCAACCAAGAGAGAAAAAGAAAG 840
Db 781 GAATTCAAAAAGACCATAGACAAAAATTAAGATGCAACCAAGAGAGAAAAAGAAAG 840
Oy 841 TTGTACACAGGCCAGTACAGCTGTCTCATATTAACAACAGCTTGAAGAGCCCATAC 900
Db 841 TTGTACACAGGCCAGTACAGCTGTCTCATATTAACAACAGCTTGAAGAGCCCATAC 900
```

```
Oy 901 CTCATCAGGCTACTGGAGAGGCGATAGACACCCCTCAAGAAATGAATAATCAAGAA 960
Db 901 CTCATCAGGCTACTGGAGAGGCGATAGACACCCCTCAAGAAATGAATAATCAAGAA 960
Oy 961 CTGCTCGACAGATTAATGAATTAAGAAATCTCCGCCAGCCCACTGGGAAACCCCT 1020
Db 961 CTGCTCGACAGATTAATGAATTAAGAAATCTCCGCCAGCCCACTGGGAAACCCCT 1020
Oy 1021 AACACGCTGCTGGACACAGACAAAGATTAAGAGACACGAGAAAGACATCAAGAGTC 1080
Db 1021 AACACGCTGCTGGACACAGACAAAGATTAAGAGAGCAGAGAAAGATCAAGAGTC 1080
Oy 1081 GCCAAACCATTAAGTTCAATAGATCTGCTTACTGATCCCTTGAGCTGAGATAC 1140
Db 1081 GCCAAACCATTAAGTTCAATAGATCTGCTTACTGATCCCTTGAGCTGAGATAC 1140
Oy 1141 TACTTGAGAGAGAAAGATTAAGATTAAGACATCTCCGCCAAAGTCGAGACAAAGAAATCA 1200
Db 1141 TACTTGAGAGAGAAAGATTAAGATTAAGACATCTCCGCCAAAGTCGAGACAAAGAAATCA 1200
Oy 1201 ACCGAACCTAATGAATATCCCAATGTGTGACGTACCTCTGTCTTATTAAGATATCAAC 1260
Db 1201 ACCGAACCTAATGAATATCCCAATGTGTGACGTACCTCTGTCTTATTAAGATATCAAC 1260
Oy 1261 AACGCTCTCAAGAGACCTCAATAGCTTGGTGATGATTAACCCCTTGATTAACGAA 1320
Db 1261 AACGCTCTCAAGAGACCTCAATAGCTTGGTGATGATTAACCCCTTGATTAACGAA 1320
Oy 1321 GAACCCCTTAAGAAATATCTACACAGACAAATGAGAGAAAGATTATCAACGAAATCAAG 1380
Db 1321 GAACCCCTTAAGAAATATCTACACAGACAAATGAGAGAAAGATTATCAACGAAATCAAG 1380
Oy 1381 GAGAAATCAAAATTAAGAAAGAAATTAAGAAATGAGAAATTAAGAAATTAAGAAATCAAG 1440
Db 1381 GAGAAATCAAAATTAAGAAAGAAATTAAGAAATGAGAAATTAAGAAATTAAGAAATCAAG 1440
Oy 1441 AGCAAAAGTCTAAAGATATCACTAAAGATTAAGAAATGAGAAATTAAGAAATTAAGAAATCAAG 1500
Db 1441 AGCAAAAGTCTAAAGATATCACTAAAGATTAAGAAATGAGAAATTAAGAAATTAAGAAATCAAG 1500
Oy 1501 TCCAAAATTCACAAATTAAGATTAAGAAATGAGAAATTAAGAAATTAAGAAATTAAGAAATCAAG 1560
Db 1501 TCCAAAATTCACAAATTAAGATTAAGAAATGAGAAATTAAGAAATTAAGAAATTAAGAAATCAAG 1560
Oy 1561 TCTTACAAAGTGGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1620
Db 1561 TCTTACAAAGTGGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1620
Oy 1621 CATATCTTGAAGAGCTGACCAAAAGCTTAAAGTATGAGAGATTAATCTCTGGGAAAC 1680
Db 1621 CATATCTTGAAGAGCTGACCAAAAGCTTAAAGTATGAGAGATTAATCTCTGGGAAAC 1680
Oy 1681 ATTGTTGTGAGAAAGAAATTAAGATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1740
Db 1681 ATTGTTGTGAGAAAGAAATTAAGATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1740
Oy 1741 ATCGAGAGCTTGTGAGAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1800
Db 1741 ATCGAGAGCTTGTGAGAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1800
Oy 1801 ACAAAGAGCAAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1860
Db 1801 ACAAAGAGCAAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1860
Oy 1861 CAAAGTGCAGAGAGTGTCTCTCATGAAACAAAGATTAAGAAATGAGAAATGAGAAATGAGAAATCAAG 1920
Db 1861 CAAAGTGCAGAGAGTGTCTCTCATGAAACAAAGATTAAGAAATGAGAAATGAGAAATGAGAAATCAAG 1920
Oy 1921 CTGAAGAAAGCTGAGATTAAGAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1980
Db 1921 CTGAAGAAAGCTGAGATTAAGAAATTAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATCAAG 1980
```



```

Db 4141 TTCCTAATGATGTCGTGGGTAATTACAGATCCAGACGAATAACAACTGACCTT 4200
QY 4201 GACTCTATTAAAAAGTATATACAGATTAAGGCGGAGTAATGAAAAATATCTGCCCTC 4260
Db 4201 GACTCTATTAAAAAGTATATACAGATTAAGGCGGAGTAATGAAAAATATCTGCCCTC 4260
QY 4261 CTGATATACATGAAACCCCTGTACAGACAGTGAACGACAAATCGACCTTCTGTAAT 4320
Db 4261 CTGATATACATGAAACCCCTGTACAGACAGTGAACGACAAATCGACCTTCTGTAAT 4320
QY 4321 CACCTGGAGGCCAAGGTCTCTACTATATCTTACGAGAAAGACATGTGGAAATTAATC 4380
Db 4321 CACCTGGAGGCCAAGGTCTCTACTATATCTTACGAGAAAGACATGTGGAAATTAATC 4380
QY 4381 AAGGAGCTGAACTACCTCAAAACATCCAAACAGCAGGAGATTTGAGAAAAATTAAC 4440
Db 4381 AAGGAGCTGAACTACCTCAAAACATCCAAACAGCAGGAGATTTGAGAAAAATTAAC 4440
QY 4441 AATTTGTCGGAATTCGAGACCTGTCTACCGATTATTAACCAACAAATCTCTGACCAAG 4500
Db 4441 AATTTGTCGGAATTCGAGACCTGTCTACCGATTATTAACCAACAAATCTCTGACCAAG 4500
QY 4501 TTTCTGTCTACCTGCGATGCTTTGCAAAACCTCGCCAAAACAGTCTGAGCAATCTGCTC 4560
Db 4501 TTTCTGTCTACCTGCGATGCTTTGCAAAACCTCGCCAAAACAGTCTGAGCAATCTGCTC 4560
QY 4561 GAGGGAACCTGCGAGGCGATGCTGAACATCTCCAGCAACCAATGGCTGAAGAAACAGTGC 4620
Db 4561 GAGGGAACCTGCGAGGCGATGCTGAACATCTCCAGCAACCAATGGCTGAAGAAACAGTGC 4620
QY 4621 CCCCAGATAGCGGCTGTTTTCAGGCATCTGAGCAGCGCAGAAAGTGAAGTGTCTCTG 4680
Db 4621 CCCCAGATAGCGGCTGTTTTCAGGCATCTGAGCAGCGCAGAAAGTGAAGTGTCTCTG 4680
QY 4681 AACTACAAACAGAGAGATTAAGTGCCTGAGAAACCCAAACCTACTCTGCAATGAACAC 4740
Db 4681 AACTACAAACAGAGAGATTAAGTGCCTGAGAAACCCAAACCTACTCTGCAATGAACAC 4740
QY 4741 AATGCGGGGTGTGACGCCGATGCTAAATGACCGAGCAAGAGACAGGGCTTAAGAGAAAG 4800
Db 4741 AATGCGGGGTGTGACGCCGATGCTAAATGACCGAGCAAGAGACAGGGCTTAAGAGAAAG 4800
QY 4801 AAAATCAGATCGAGTGTACTTAAGCCGACATCTATCCACTCTTGCAGCGGATTTTTCG 4860
Db 4801 AAAATCAGATCGAGTGTACTTAAGCCGACATCTATCCACTCTTGCAGCGGATTTTTCG 4860
QY 4861 TCCAGCTCTAATTTCTCGGCGATCTCTTCTGCTGATCTCTCATGCTGATCTCTGACGC 4920
Db 4861 TCCAGCTCTAATTTCTCGGCGATCTCTTCTGCTGATCTCTCATGCTGATCTCTGACGC 4920
QY 4921 TTTCACTAATGATCGATGG 4940
Db 4921 TTTCACTAATGATCGATGG 4940

```

RESULT 2
AAV35363
ID AAV35363 standard; DNA; 4940 BP.

XX AAV35363;
AC
XX
XX 23-SEP-1998 (first entry)
XX
XX P. falciparum gp190 DNA.
XX
XX gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX monoclonal antibody; passive immunisation; parasite; ss.
XX
XX Plasmodium falciparum.
FH Key Location/Qualifiers
FT CDS 10..4929
FT /*tag= a

```

FT /product= gp190
XX
XX MO9814583-A2.
XX
XX 09-APR-1998.
XX
XX 02-OCT-1997; 97WO-EP05441.
XX
XX 02-OCT-1996; 96DE-4040817.
XX
XX (BUA/) BUJARD H.
XX
XX Bujard H, Pan W, Tolle R;
XX
XX WPI: 1998-240088/21.
XX
XX P-PSDB: AAM54145.
XX
XX Recombinant production of complete gp190/MSP-1 plasmodium surface
XX protein - useful in anti-malaria vaccines, also stabilising genes by
XX reducing their AT content
XX
XX Example 1: Fig 3c; 48bp; German.
XX
XX This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
XX surface protein). This gene is used in a method for stabilising the
XX gene sequences by reducing the AT content. Such products are useful in
XX vaccines against malaria or for producing monoclonal antibodies (for
XX passive immunisation). The complete gp190 protein can now be produced
XX outside the parasite and has, at least over extended regions, the native
XX pattern of folding. Larger amounts of the protein can be produced
XX recombinantly than would be possible using the parasites as source.
XX
XX Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other;
XX
Query Match 57.4%; Score 2837.6; DB 19; Length 4940;
Best local similarity 73.4%; Pred. No. 0;
Matches 3626; Conservative 0; Mismatches 1314; Indels 0; Gaps 0;
QY 1 CGACGCGCTAAGAAATATTTCTCTCTGTCATCTTCTGTTTATCATCAATACT 60
Db 1 CGACGCGCTAAGAAATATTTCTCTCTGTCATCTTCTGTTTATTAATAATCA 60
QY 61 CAGTGCCTGACCCAGAAATCCATCAGAGCTGTTAAGAAACCTGGAAGAT 120
Db 61 CAATGTGATACACATGAAGATTAATCAAGACTTGTCAAAAACATGAGAGCTTGAAGAT 120
QY 121 GCCGTCTTACCGGATACAGCCTGTTCAGAAAGAGAAATGTGCTGAATGAAGGACG 180
Db 121 GCAGTATTGACAGGTTATGATTATTTCAAAAAGAAAATGTATTAATGAAGGACCA 180
QY 181 AGTGCACAGGCGCTTACACACGACACACCGGTTCTAAAGGCTGTGGCTAGCGGTGC 240
Db 181 AGTGCACAGGCTGTACACAGTACAGTACCTGTCAAGGCTTCAAGGCTTCAAGGTGT 240
QY 241 TCCGCTGGGTCTGTGCGCTCTGCGGCTTCCGCTCGCCTCGGCGGAGCGTGGCATCAGT 300
Db 241 TCAGGTGCTCAGTTGCTTCAAGTGTGCTCACTGCTTCAAGTGTGCTCAGT 300
QY 301 GCGTCACTGAGGAGCGGCTTCCGGGACAGTGCAGAACCAATCACTTGCACACTCT 360
Db 301 GCGTCACTGAGGAGCGGCTTCCGGGACAGTGCAGAACCAATCACTTGCACACTCT 360
QY 361 AGCGATTCGAGCGCAAGTCTTACGCGCAGCTCAAGCAACGAGTGAAGAAATATCTCTC 420
Db 361 AGTATTCAGATGCTTAAATCTTACGCTGATTTAAACACAGATGCAAGAAATTAATCTG 420
QY 421 ACTATCAAGAGAGCTGAAGTATCCACAGTGTTCGACCTCACTAATCATATCTCTGACACTG 480
Db 421 ACTATCAAGAGAGCTGAAGTATCCACAGTGTTCGACCTCACTAATCATATCTCTGACACTG 480
QY 481 TGTGATTAACATTCATGCTTCAATATCTGATGTGACGGTTCAGAGAGATCAATGAAGTCTC 540
Db 481 TGTGATTAATTAATCATGTTCAAAATATTAATGTGATGATGAAGAAATTAATGAATTA 540

```



```
QY 541 CTGTACAGTGAATTTCTACTTCGACTTCGTAAGGGCCAACTGATGACTTTGGCC 600
  || || || || || || || || || || || || || || || || || || || || ||
Db 541 TTATATAAATTAACTTTATTTGATTTATTAAGACAAATTAATGATGATGCT 600
QY 601 AATGACTATTGTCAAATTCATTCAATTTGAGATCAGAGCCCAAGATGGACGATGG 660
  || || || || || || || || || || || || || || || || || || || || ||
Db 601 AATGATTAATGTCAAATTCATTCAATTTGAGATCAGAGCCCAAGATGGACGATGG 660
QY 661 AAGAGTGGTCTTGGATTCGACAGCCCTCGACACATCAAGGACATGGGGAAAG 720
  || || || || || || || || || || || || || || || || || || || || ||
Db 661 AAAAACTGTGTTCGATTTAGAAACCATTTGACATATTTAAGATTAATGAGAAA 720
QY 721 ATGGAAGATTTATTTAAAAAAGATTAAGAACCATCGAGAACATTAAGCATGATGAA 780
  || || || || || || || || || || || || || || || || || || || || ||
Db 721 ATGGAAGATTTATTTAAAAAATTTAAAAACCATAGAAAATTAATGATTAATGAA 780
QY 781 GAATCCAAAAAGCCATAGACAAAATTAAGATTAAGCAACCAAGAGAGAAAGAAAG 840
  || || || || || || || || || || || || || || || || || || || || ||
Db 781 GAAAGTAGAAGAAACAAATGATTAAGAAATTAAGAACTTAAGAGAAAGAAAAA 840
QY 841 TTGTACAGGCGCAGTACGACCTGTCCATCTATACAAACAGCTTGAAGAGCCATAC 900
  || || || || || || || || || || || || || || || || || || || || ||
Db 841 TTATACCAAGCTCAATATGATCTTCTATTTACAAATTAAGCAATTAAGAGACATAT 900
QY 901 CTCATCAGGCTACTGGAGAGCGCATAGACCCCTCAAGAGATGAAAATATCAAGAA 960
  || || || || || || || || || || || || || || || || || || || || ||
Db 901 TTATATAGCGCTTTAGAAAACCTATGACCTTTAAAAAATGAAAACATTAAGGAA 960
QY 961 CTGCTGACAGATTTATAGAAATTAAGAAATCTCGGCGACCAACTCGGACACCCCT 1020
  || || || || || || || || || || || || || || || || || || || || ||
Db 961 TTACTGATAGATTAATGAAATTAAGAAATCCCGCCACGCGCAATCTGGAATATACCA 1020
QY 1021 AACACGCTGTGACAGAAAGAAAGATAGAGAGACGAGAAAGAGATCAAGAGATC 1080
  || || || || || || || || || || || || || || || || || || || || ||
Db 1021 AATACCTCTCTGATTAAGAAACAAAAATAGAGAGACGAGAAAGAAAGAAAT 1080
QY 1081 GCCAAACCATTTAAGTTCAACATAGATCTCTCTTTACTGATCCCTTGAGCTGAGTAC 1140
  || || || || || || || || || || || || || || || || || || || || ||
Db 1081 GCCAAACCATTTAATTTAATTTAGTTAGTTTACTGATCCCTTGATTAATGAGATAC 1140
QY 1141 TACTTGAAGAGAAAGATTAAGATTAAGATCTCGGCCAAAGCTGAGAAAGGATCA 1200
  || || || || || || || || || || || || || || || || || || || || ||
Db 1141 TATTTAAGAGAAAAATTAATTAATTTGATTAATGATGCAAAAGGTTGAAACAAAGGATCA 1200
QY 1201 ACCGAACCTAATGAAATATCCCAATGCTGACGCTGCTGCTTATAACGATATCAC 1260
  || || || || || || || || || || || || || || || || || || || || ||
Db 1201 ACTGAACCCCAATGAAATATCCCAATGCTGACGCTGCTGCTTATAACGATATTAAC 1260
QY 1261 AAGCTCTCAGAGAGCTCAATAGCTTCGCTGCTGATTAACCCCTTCGATTTAGGAA 1320
  || || || || || || || || || || || || || || || || || || || || ||
Db 1261 AATGCTTTAATGAACTTAATCTTTGCTGATTAATTAATCAATTTGATTAACAAA 1320
QY 1321 GAACCCCTAAGAAATTTCTACAGACAAATGAGAGAAAGATTTTCAACGAATCAAG 1380
  || || || || || || || || || || || || || || || || || || || || ||
Db 1321 GAACCAAGTAAACATTAATGATGATTAATGAGAGAAAGATTTCAATTAATGAAATTAAG 1380
QY 1381 GAGAGATCAAAATTTGAGAGAAAGAAATTTGAGAGTACAGAAAGATTTGAGAGACCG 1440
  || || || || || || || || || || || || || || || || || || || || ||
Db 1381 GAAAAAATTTAAATGAGAAAAAATTTGAGATCTGATTAATAAATCTTGAGAGAGAG 1440
QY 1441 AGCAAAAGTCTAAAGATATCACTAAAGAGTATGAAAAGCTGCTGAGACGATCTATGAT 1500
  || || || || || || || || || || || || || || || || || || || || ||
Db 1441 TCTAAGCTCTTAATGATTAATCAAAAGAAATGAGAAAAATTTACTTATGAAATTTATGAT 1500
QY 1501 TCCAAATTCAGACATTAACATCGACCTGACCAACTTGAGAGAAATGATGGAAAAAGGTAC 1560
  || || || || || || || || || || || || || || || || || || || || ||
Db 1501 AGCAATTTCAATATATATATATGATTTAAGTAAATTTGAGAAAAATGATGGTAAAGATAT 1560
QY 1561 TCTTACAAAGTGAAGAACTGACACACATTAATACCTTTGATGATGATGATGATGATGAT 1620
  || || || || || || || || || || || || || || || || || || || || ||
Db 1561 TCAATATAAGTTGAGAAACTTACACACATTAATTAATTTTGGATCTGATGAAAAATTTCTAAA 1620

QY 1621 CATTAATCTTGAGAGAGCTCACCAAGCTCTTAAGTATATGAGAGACATTAATCTGCGGAC 1680
  || || || || || || || || || || || || || || || || || || || || ||
Db 1621 CATTAATCTTGAGAGAGCTCAATTAAGGCTCTTAATTAATGAGAGATTAATCTTTAAGAAAT 1680
QY 1681 ATTGTTGTGAGAGAAAGACATTAAGTATTAAGAAATCTCTAAGTAAAGTGAAGAGAG 1740
  || || || || || || || || || || || || || || || || || || || || ||
Db 1681 ATAGTGTGAGAGAAAGACATTAAGTATTAAGAAATTTAAGCAAAATGAGAAATGAG 1740
QY 1741 ATGAGAGCGCTGTGAGACATTAAGAAAGATGAAGAACAGTGTGTTGAGAGAAAGAT 1800
  || || || || || || || || || || || || || || || || || || || || ||
Db 1741 ATGGAACATTAATGAGAAATTTAAGAAAGATGAAGAACAGCTTTTGAAGAAAAAT 1800
QY 1801 ACAAAAGACGAAATTAATTAACGATGAGAGATCTGAGAGCTTCGATATTTGAAAGTC 1860
  || || || || || || || || || || || || || || || || || || || || ||
Db 1801 ACTTAAGACGAAATTAATTAACGATGAGAGATTTTGAAGATTAATGATTAAGAAAGT 1860
QY 1861 CAAGTGAAGAGGTGCTCCATGAACAAAGATTAAGAAATCAAGAAAGTCAAGTAT 1920
  || || || || || || || || || || || || || || || || || || || || ||
Db 1861 CAAGTGAAGAGGTGCTTATTAATGAACAAAGATTAAGAAATTAAGAAAGTCAAGTAT 1920
QY 1921 CTGAGAGAGCTGAGATTAAGAAATTAATTAATGATGCGGAATGTTAAGAGAGAAAT 1980
  || || || || || || || || || || || || || || || || || || || || ||
Db 1921 TTAAGAAATGATGAATTAAGAAATTAATTAATGATGCGGAATGTTAAGAAAGAAAT 1980
QY 1981 AAGCAGAGACCTACTATCTCATCTGATCTGATCAAGAAAGATTAAGCAAACTGAAAGTTC 2040
  || || || || || || || || || || || || || || || || || || || || ||
Db 1981 AAGCAAGAAACCTTATTAATTAATGATGTTGAAAAAAGTTATTAATTAATTAAGTTC 2040
QY 2041 ATGCCCAAGTGCAGAGCCGATCAAGAGAGAGAGAAAGAAAGATTAAGTGAAGAGACAG 2100
  || || || || || || || || || || || || || || || || || || || || ||
Db 2041 ATGCCCTAAGTAAATATCATTTGATTAATTAAGAAAGAAAGAAAGATTAAGTGAAGAGACAG 2100
QY 2101 TCAGATTAATCTCGAGCCCTTCAGAGAGAGAGAGATTAAGCCGAGAGCTACACCAAGCC 2160
  || || || || || || || || || || || || || || || || || || || || ||
Db 2101 TCGGATTAATCTCGAGAGCCCTTCAGAGAGAGAGAGATTAAGCCGAGAGCTACACCAAGCC 2160
QY 2161 GCAACAGAGCCGCTTACAGCTCGAAGAGAGATGAGTGAAGCTGCAAGCAAGCAAGAGAG 2220
  || || || || || || || || || || || || || || || || || || || || ||
Db 2161 GCAACAGAGCCGCTTACAGCTCGAAGAGAGATGAGTGAAGCTGCAAGCAAGCAAGAGAG 2220
QY 2221 AACAGAGCAGAGCCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
  || || || || || || || || || || || || || || || || || || || || ||
Db 2221 AACAGAGCAGAGCCCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 CCACAGAGCTCTGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
  || || || || || || || || || || || || || || || || || || || || ||
Db 2281 CCACAGAGAGCTCTGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2341 CTGTATGAGTCTGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
  || || || || || || || || || || || || || || || || || || || || ||
Db 2341 CTGTATGAGTCTGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
QY 2401 ACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
  || || || || || || || || || || || || || || || || || || || || ||
Db 2401 ACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
QY 2461 TCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
  || || || || || || || || || || || || || || || || || || || || ||
Db 2461 AGTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
QY 2521 TCTATGTTGATAGCTCAACAAATTTCTCTCTCAAGCTGTTCAAGATATATGAGAGAG 2580
  || || || || || || || || || || || || || || || || || || || || ||
Db 2521 TCTATGTTGATAGCTCAACAAATTTCTCTCTCAAGCTGTTCAAGATATATGAGAGAGAG 2580
QY 2581 GAGATGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
  || || || || || || || || || || || || || || || || || || || || ||
Db 2581 GAGATGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
QY 2641 GAAGCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
  || || || || || || || || || || || || || || || || || || || || ||
Db 2641 GAAGCGAAAAAGATTAATCAACATCTGTAAGAAAGCTTTTCAAGTATCAATGAGAGAGATTA 2700
QY 2701 TCTCTGACAGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
  || || || || || || || || || || || || || || || || || || || || ||
```



```
Db 2701 TCATTAAACCTCGAGATAAACCAGTAAGTGCATAATGATACATCTCTACA 2760
Qy 2761 AACCTTAATACTACGTAACCTGTTTGAGAACATCTGCTCTCGCAGAAATAGAAC 2820
Db 2761 AATTGAATATAGTGTAAATTTATTTGAAAACAATCTGAGCTTGGAAAAACAAT 2820
Qy 2821 ATCTACCAAGAACTTATGACAGAAATCGTCCGAACCTTCTACGAAGAATACTGAAA 2880
Db 2821 ATATACCAAGAAATTAATAGTGGTCAAAAAAGTAGTGAACCTTTATGAAAAGTATATAAA 2880
Qy 2881 GACACGACACATCTCTATTAAGAGAGAGCTTCACTAATCTGTAATCTAAAGCCGAT 2940
Db 2881 GATAGTGTATACATTTTATATGAATCTTTACAAATTTTGTAAATCTAAAGCTATGAT 2940
Qy 2941 ATCAACTCTTAAAGATGAATCTAAACGTAAGAGCTGAAGAGACATCAATAGCTG 3000
Db 2941 ATTAATTCATTTGAATGAATGAAATCAAAAAGAAATTAGAAAGATATTTAATAATTA 3000
Qy 3001 AAGAAACACTGCAACTGACCTTGACCTGTACAAACAGTCAAACTGAACCTGAGAGA 3060
Db 3001 AAAAAAATTTACAGTATATCAATTTGATTTATTAATTAATTAATTAATTAAGAAA 3060
Qy 3061 CTCTTCGACAAGAAAGACAGTCCGCAAGTATAGATGACATCAAGAGTTCAGCTG 3120
Db 3061 TTATTTGATTAAGAAAGAAAGAGTGTGTAATATATAATGAATTAAGAAACTTACTTTA 3120
Qy 3121 CTCAGAGGACAGCTGTAAGCAAACTCACTGAACTGAAATCGAAACAGTACTGAG 3180
Db 3121 TTAAAGAAACATTAAGATTAATAATTTGAATTAATTAATTAATTAATTAATTA 3180
Qy 3181 AACTTCAGTGTCTTCAACAAGAAAGAAAGCCGAGATCGCCGACACAGAAACCT 3240
Db 3181 AACTTCCTGTTTCTTTAAACAAAAAAGAAAGCTGAATGACGAAGCTGAACACACA 3240
Qy 3241 CTGAGAAACCAAGATTCCTCTCAAACTACAAAGCCCTGCTCAAGTATTAATGAGC 3300
Db 3241 TTAGAAAACCAAAAATTTATTTGAACATTAATAAGACCTGTAATTAATTAATGAT 3300
Qy 3301 GAGCTTCTCTCTGAGAGACTCTCTCCGAGAGAGCATCCAGCCGAGATTAACGCC 3360
Db 3301 GAATCATCTCATTAATAAAATTTAGTGAATCAATTAATTAATTAATTAATTAATTA 3360
Qy 3361 AGCCTCGAAGACTTCAGGCTCTGCTAAGCTCGAAAGCAGCTGAGAGCAACTGAGC 3420
Db 3361 AGTTTAGAAAACTTTAAAGTATTAAGTAAATTAGAAGAAAAATTAAGATTAATTAAT 3420
Qy 3421 CTGAGAAAGAAAGCTAGCTACCTCTCTAGCGAGCTGATCAACTGATCGCCGAGCTC 3480
Db 3421 TTAGAAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3480
Qy 3481 AAGGAAGTCAATTAAGAAACAAAGTACACCGCAATAGCCCAAGCGAATTAATACAGC 3540
Db 3481 AAGGAAGTCAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3540
Qy 3541 GTGATAACGCACTGAGATTTTACAAAGTTCCTGCTGGAAGAAAGATGTGCCACT 3600
Db 3541 GTTAACAATGATTAAGATTTTACAAAAATTTCTCCGAGAGAAAGATGTGTCAACA 3600
Qy 3601 GTGTGTCGATGTGCTGCTCGACACACTGAGAGTCAACCTAAGAACGCTGCATCT 3660
Db 3601 GTTGAAGTGAAGTGAATGATCGACACATTTAGAACAAAGTCAACAAAGAACACATCA 3660
Qy 3661 ACTCATGTCGAGCGAGATCCATTAATCAATTAACACATCTCAGAACGTGAGATAGAGTC 3720
Db 3661 ACTCATGTCGAGCGAGATCCATTAATCAACATTAACCAATCAACAAAGTGAATTAATTA 3720
Qy 3721 GATGACGTCATCATTTGCTGCTGAGAGAGAGAGAGAGATGATGATGAGCTCGGC 3780
Db 3721 GATGACGTCATCATTTGCTGCTGAGAGAGAGAGAGAGATGATGATGATGAGCTCGGC 3780
Qy 3781 CAGGTGTCACCGGTGAGGCTGTCACCTCTCCGATTTGATTAACATTTCTGCAAAATC 3840
Db 3781 CAGGTGTCACCGGTGAGGCTGTCACCTCTCCGATTTGATTTTCTGCAAAATC 3840
```

```
Db 3781 CAAGTAGTAACGAGAGAAAGCACTACCTCTCCGTAATTTGATTAACATCTCTTAAT 3840
Qy 3841 GAGAAAGATACGAAGTCTCTATCTGAACCTCTGCGAGCGGTATAGGCTCTCAG 3900
Db 3841 GAAAAATGATATGAGGTTTTATATTTAAACCTTTAGCAGGCTTTATAGAAATTAATA 3900
Qy 3901 AAACAGCTGGAATTAAGCTGTATGACCTTCAATGTCACGTCGAAGATCTGCAAGC 3960
Db 3901 AAACATTTAGAAATTAAGCTGTATGACATTTAATGTTAATGTTAAGATATTTAATTA 3960
Qy 3961 CGCTTAATTAAGAGAGAAATTTCAAGAACGCTTGAGAGAGCAGTATTCCTATAAA 4020
Db 3961 CGATTTAATTAAGAGAGAAATTTCAAAATGTTTGAAGATCAGATTAATTCATATAAA 4020
Qy 4021 GACCTGACCTCTCTAATCTAGCTTCAAGAACCATTCAGATTCCTAATTAAGAGAG 4080
Db 4021 GATTTAATCATCAAGTAATTAATGCTGCAAGATCCATTAATTTCTTAATTAAGAAAA 4080
Qy 4081 AGGATTAATTTCTGCTAGTTTACAACTATATCAAGGACTCCATGACACCGATTAAT 4140
Db 4081 AGAGATTAATTTCTTAGCAGTTATATTAATTAATTAAGATTCATTAATGATGAGATTAAT 4140
Qy 4141 TTGCTAATGATGTGCTGGGTATTTACAAAGATCTGACGGAATAATTAACAGTTCAGCTT 4200
Db 4141 TTGCAATTAATGATGTGCTGGGTATTTACAAAGATCTGACGGAATAATTAATTAATTA 4200
Qy 4201 GACTTATTAATTAATGATTAATCAAGATTAAGCAAGCGGATGAATAATTCGCTTC 4260
Db 4201 GATTCATTAATTAATTAATTAATCAAGCAAGCAAGCAAGGATGAATAATTAATTCCTC 4260
Qy 4261 CTGAATTAATCAAGAACCTGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 4320
Db 4261 TTAACCAATTTAGACACTTATATTAATTAATTAATTAATTAATTAATTAATTAATTA 4320
Qy 4321 CACCTGAGAGCCAAAGTCTCAACTATCTTAACGAAGAGCAATGTAAGTAAATC 4380
Db 4321 CATTAGAGCAAAAGTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4380
Qy 4381 AAGAGCTGAACTACTCAAAATTAATTAAGCAAGCGGAGATTTCAAGAAAAATTAAC 4440
Db 4381 AAGAGCTGAACTACTCAAAATTAATTAAGCAAGCGGAGATTTCAAGAAAAATTAAC 4440
Qy 4441 AATTTGCTGCAATGTCAGACCTGCTACGATTAATTAACCAACAACTCTGCTCAAG 4500
Db 4441 AATTTGCTGCAATGTCAGACCTGCTACGATTAATTAACCAACAACTCTGCTCAAG 4500
Qy 4501 TTTCTGCTCAGTGGCATGCTGTTGAAAACCTGCGCAAAACAGTCTGAGCAATCTCTC 4560
Db 4501 TTCTTATTAAGAGATGATGATGTTTGAATAATCTTGCTAAACCGTGTATTAATTAAT 4560
Qy 4561 GACGCAACCTGCAAGGCTGTAACATCTCCAGCAACCAATGCGTGAAGAAACAGTGC 4620
Db 4561 GATGGAACCTGCAAGGCTGTAACATCTCCAGCAACCAATGCGTGAAGAAACAGTGC 4620
Qy 4621 CCCGAGATGAGGCTGTTTCAAGCATGTCGAGAGCGCCAGAGATGCAAGTCTGCTG 4680
Db 4621 CCAGAAATTTCTGATGTTTCAAGCATTTAATGAAGAAAGAAATTAATTAATTAATTA 4680
Qy 4681 AACTACAAACAGAAAGATTAAGTGTGCGTGAAGAAACCAAACTCTGCAATGAAGAAC 4740
Db 4681 AATTCAAAACAGAAAGATTAAGTGTGTTGAAAATCCAAATCTCTGTAACCAAAAT 4740
Qy 4741 AATGCGGCTGTCAGCGGATGCTAATTAAGCAACCGGAGAAAGCAAGCGCTCTAAGGAAG 4800
Db 4741 AATGCGGCTGTCAGCGGATGCTAATTAAGCAACCGGAGAAAGCAAGCGCTCTAAGGAAG 4800
Qy 4801 AAAATCAACATGAGTACTTAAGCCGACATCTTAATCCACTTTCGAGGAGATTTTTCG 4860
Db 4801 AAAATCAACATGAGTACTTAAGCCGACATCTTAATCCACTTTCGAGGATTTTTCG 4860
Qy 4861 TCCAGCTCTAATTTCTGCGGATCTCTCTGCTGATGCTCAATGCTGATCTCTGACAGC 4920
Db 4861 AGTTCTCACTCTTATGAGATATCATCTTATTAATTAATCAATGATTAATTAATCACT 4920
```


Oy	4921	TTTCATCTAATAGATGATGG	4940
Dd	4921	TTCAATTATAAGATGCATGG	4940
RESULT 3			
ID	AAN50530	standard; DNA; 5760 BP.	
xx	AAN50530;		
AC	AAN50530;		
XX			
DT	30-SEP-1991	(first entry)	
XX			
DE	Sequence encoding the P195 protein of Plasmodium falciparum.		
XX			
KW	Malaria vaccine; epitope; antigen; immunogen; ss.		
OS	Plasmodium falciparum.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	216..5179	
FT		/tag= a	
PN	EP154454-A.		
XX			
PD	11-SEP-1985.		
XX			
PE	21-FEB-1985;	85EP-0301173.	
XX			
PR	26-SEP-1984;	84GB-0024340.	
XX			
PR	22-FEB-1984;	84GB-0004692.	
XX			
PR	21-FEB-1985;	85GB-0004429.	
PA	(WELL) WELLCOME FOUNDATION LTD.		
XX			
PI	Holider A,	Sandhu J,	Odink K,
XX			Lockyer M,
DR			Riveros-Moreno V;
XX			
WI	P-I: 1985-224845/37.		
DR	P-PSDB; AAP50777.		
XX			
PT	Cloned DNA sequence encoding plasmodium falciparum protein -		
PT	useful for expressing the protein for use in vaccines against		
PT	malaria		
PS	Claim 2; Fig 1; 51pp; English.		
CC			
CC	The sequence encoding the P195 protein of Plasmodium falciparum		
CC	(AAN50530) and a peptide comprising at least one of its epitopes		
CC	(see AAP50777) are claimed. Also claimed is a vaccine for inducing		
CC	immunity to malaria comprising the novel peptide or P195 or a		
CC	peptide comprising at least one epitope when derived from the new		
CC	DNA sequence, together with a carrier.		
XX			
SEQ	Sequence 5760 BP; 2565 A; 630 C; 725 G; 1840 T; 0 other;		
Query Match	54.8%;	Score 2706.2;	DB 6; Length 5760;
Best Local Similarity	72.1%;	Pred. No. 0;	
Matches 3587; Conservative	0;	Mismatches 1343;	Indels 45; Gaps 3;
Oy	10	ATGAAATATATTTTCTCTCCTCGTGTCATTTCTGTTTTATCATCAATACTAGTGGCG	69
Dd	216	ATGAAGATCATATTTCTTTTATGTTCATTTCTTTTATATATAATACACAATGTGA	275
Oy	70	ACCACGAATCTTATCAGAGAGCTGGTTAGAAGCTGTAAGAGCTTTGGAAATGCCGCTT	129
Dd	276	ACACATGAAGATTATCAGAACTTGTCAAAAACTGAAAGCTTTAGACATGTCAGATTG	335
Oy	130	ACCGATACAGCCTGTTCCAGAAGAGATGTCCTAATGAAGAGCAGATGGCAG	189
Dd	336	ACAGGTTATAGTTTATTTCAAAGAAAATGGTATTAATGAAGAACAGATGGAC	395
Oy	190	GCGGTACACCAAGCACACCCGGTTTAAAGGCTGTGGCTAAGCGGTGGCTCCGGTGG	249

Db	396	GCCTGTACACCTACACCTGGTTCAAAAGGGTTACGTTTCAGTGGGTTCAAGGTGC	455
Qy	250	TCCTGGCCCTCTGGGGGTTCCGTCGCCCTCCGGCGCAGCGTGGCATCAGGTGGCTCA	309
Db	456	TCAGTTCGCTCAGGTGGCTCAGTTGGCTTCAGGGGCTCGAGTTCAGGTGGCTCAG	515
Qy	310	GCAAGCGCGGTTCCGGGACAGTCGAGAACCAATCCATCTGACACCTAGACATCC	369
Db	516	GCCTCAGGTGGTTAGGTAAATTTCAGACGCTACAAATCTTCAGATTAATTCAGATG	575
Qy	370	GACGCCAAGTCTACGCCGACCTCAAGCACCGAGTGAAGAACTATCTCCTACTCA	429
Db	576	GATGCTAAATCTTCGCTGGATTATAAACAAGAGTACGAATTAATCTGTTAACTCA	635
Qy	430	CAGCTGAAGTACCCACAGTTGTCGACCTCATACTATCATATGCTGACACTGTGTAT	489
Db	636	GAACCTCAATATCCTCAACCTTTGGATTAACTAACTAATGTTAACCTTGTGTATAT	695
Qy	490	ATTCATGCTTCAATATCTGATTGATGGGTTACGAGAGATCAATGAACCTCCTGTCA	549
Db	686	ATTCATGCTTCAATATTTTAATTTGATGATATGAAGAAATTAATGAATTAATATAA	755
Qy	550	TTGAATTTCTACTTCGACTGCTAAGGGCCAACTGAATGACGTTTGGCCAACTGAT	609
Db	756	TTAAACCTTTATTTTGATTATTAAGAAGCAAAATTAATGATGTAATGCTAATGAT	815
Qy	610	TGTCAAATTCATTCAATTGGAAGATCAGAGCCACAGAGTTGGACGCTATGGAAGA	669
Db	816	TGTCAAATACCTTTCATCTTAAATTCGTCGCAAAAGAAATTTAGACGTAATTA	875
Qy	670	GTCCTCGATATGCGCAAGCCCTCGACCAATCAAGCAATGAGGAGAAATGGAAGAT	729
Db	876	GTTTCGGATATGAGAAACCTTAAGCAATATTTAAAGATTAATGTAGAAAAATGCA	935
Qy	730	TATATTAAAAAGATTAAGAAGACCATCGAATTAACGAGCTGATCGAAGATCCAA	789
Db	936	TACATTAATAAAAAATTAATAAACCATTAAGAAATTAATGAATTAATGAAGAA	995
Qy	790	AAGCATATGACAAATAATTAAGATGCAACCAAGAGAGAAAGAGAGAGTTGACAG	849
Db	996	AAAACAAATGATTAATAAATAAGAAATGCACTAAAGAAAGAAAGAAATAATTA	1055
Qy	850	GCCCGATCGACGCTGCTCATCTAATCAACAACTGGAAGAGCCATTAACCTCA	909
Db	1056	GCTCATATGATCTTCTTATTTACATTAATCAATTAAGAAAGACCATTAATTA	1115
Qy	910	GTACGAGAGCGCATATGACACCCCTAGAGAAATGAATAATCAAGAACTGCTGAC	969
Db	1116	GTTTATAGAAAAAGCTATTTGACACTTTTAAAAAAAATGAAACCTTAAGSAAT	1175
Qy	970	AAGATTAAATTAAGAAATCCTCGCGCAGCCACTGTGGAGACACCCCTAACGCTG	1029
Db	1176	AAGATTAATGAATTAATAAATCCGCCACCGCAATTCGTGAAATACCAAAATCT	1235
Qy	1030	CTGAGCAAGAACGAAGATTAAGAGAGGACAGCAAGAAAGATCAAAAGATGCCAA	1089
Db	1236	CTTGATTAAGCAAAAAATCGAGGAACAGCAAAAAAGAAATTAAGAAATTTGCC	1295
Qy	1090	ATTAGTCAACATAGAAATCTCTCTTACTGATCCCTTGAGCTGGAGTACTTCTG	1149
Db	1296	ATTAAATTAATATTTGATAGTTTATTTACTGATCCACTTGAAATTAAGATTA	1355
Qy	1150	GAGAAATTAAGAAATATAGACATCTCCGCAAGTCGAGCAAAAGAAATCAACG	1209
Db	1356	GAATAAATAAATAATTTGATATTAAGTCAAGAGTGAAGAAAGAAATCACTGA	1415
Qy	1210	AATGAATATCCCATGCTGTGAGCTACCCCTCTCTTAATAGAGATTAACAACCT	1269
Db	1416	AATGAATATCCAAATGAGATTAATCTTCTTGTATTAAGCAATTAATTAACCT	1475
Qy	1270	AACGAGTCATATCTTCGGGATGATTAATCAACCTTGATTAATCAAGAAACCT	1329

Db 1476 AATGAACCTTAATCTTTGGTGATTTAAATTAATCCATTGATTTATACAAAAGACCAAGT 1535
Qy 1330 AAGAAATCTACACAGAAATGAGAGAAAGATTATACAGGAATCAAGAGAAAGATC 1389
Db 1536 AAAAAATATATAGATTAAGAGAAAAAAATTCATTAAGAAATTAAGAAAAAAT 1595
Qy 1390 AAAATGGAGAGAAATTTAGAGATGACAGAAAGATTCAGAACCCGCGCAAAAGT 1449
Db 1596 AAAATGAAAAAATTTGATCTGATAAAAATCTTACGAAAGACAGATCTAAGTCT 1655
Qy 1450 CTAAAGATATCATTAAGAGATGAAAAAGCTGTGAACGAGATCTATGATCCAAATTC 1509
Db 1656 TTAATGATATTAACAAAGATATGAAAAATTAATTAAGAAATTTATGATAGCAAAATTC 1715
Qy 1510 AACAAATACATGACCTGACCACTGACAGAAATGATGGGAAAAAGGACTCTTAACA 1569
Db 1716 AATTAATTAATAGATTTAACTAATTTCCAAAAATGATGGGTAAGAGATTTCAATTA 1775
Qy 1570 GTGAGAAACTGACACACCATTAATACCTTTGATCTATGAGAAATTTCAAGCATATCT 1629
Db 1776 GTTGAGAAACTTACACACCTTAATCTTTGATCTCTATGAGAAATTTCAAGCATATCT 1835
Qy 1630 GAGAAAGCTACCAAGCTCTTAATATATGAGAGCTATCTCTCGGACATTTGTGTG 1689
Db 1836 GAAAAATTACAAAAGCTTTAAATATATGAGAGATTTATCTTTAAGGAATATAGTAGTT 1895
Qy 1690 GAGAAAGAACTAATATTAACAAAGATCTAATAGATGGAAGAGAGATGAGAG 1749
Db 1896 GAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1955
Qy 1750 CTTGTTGAGAACTTTAAGAGATGAGAAAGCTTTGTTGAGAGAAAGATTCAAAAGAC 1809
Db 1956 TTAGTTGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2015
Qy 1810 GAAATTAACAGATGAGAAAGATCTGAGAGCTCCGATATTTGTTAAAGTCCAAAGTCA 1869
Db 2016 GAAAAATTAACAGATGAGAAAGATTTTAAGAGATCTGACATTTGTAAGTAAGTTC 2075
Qy 1870 AAGGTCTCTCATGAGAAAGATTTGATGAACTCAAGAGACTCACTCATTTGTAAGAAC 1929
Db 2076 AAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2135
Qy 1930 GTGAGATTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1989
Db 2136 GTAGATTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2195
Qy 1990 CCATCTACCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2049
Db 2196 CCTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2255
Qy 2050 GTGAGAGCTGATCAAGAGAGAGAAAGATTAATTAATTAATTAATTAATTAATTAAT 2109
Db 2256 GTAGAAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2315
Qy 2110 TCCGAGCTTCCACAGAGAGAGATTAACCGAGAGGCTTACCCAGAGCCGCGACAGAG 2169
Db 2316 TCGGAACCATCAACCGAGAGAGAAATTAACAGCAACGACATCAAAACCTGACACAA 2375
Qy 2170 GCCGGTTAGCTCTGAGAGGAGATAGCTGCAAGCTCAAGCAAGAGAGAGAGAGCA 2229
Db 2376 GCAGAGATCTGTTAAGAGAGAGATTAAGTCAAGCAAGAGAGAGAGAGAGAGAGCA 2435
Qy 2230 CAGCCTCAGTGCAGTCCCGTTCCAGAGCTAAAGCTCAAGTCTCAACACACACACT 2289
Db 2436 CACCCACAGATCAAGTACAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2495
Qy 2290 CCTGTGAATTAACAGAGAGAGATGCAAACTGACTCACTGATGAGAGAGCTATGAG 2349
Db 2496 CCAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2555
Qy 2350 TTCCGATTAATCTCTATCTGACAAATATATCTCTGCTCTGACAGAGAGATGAGAC 2409
Db 2556 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2615

Qy 2410 GGAGAGATTCTTAACAGTACAGATTAACCAAGAGAGAGAGATTAACCTGCTCTGT 2469
Db 2616 GAAAAATTAATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2675
Qy 2470 GATCCACTGAGACCTGCTCTCAATATCCAGAAACAACTCCGTTATGATTAATGTTTC 2529
Db 2676 GATCCATTAAGATTTATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2735
Qy 2530 GATAGCTCAACA-----ATTCT 2547
Db 2736 GATAGCTTAACATAGTTTATCAACTATTTATGTTTATGAAAAAATTTGTTAT 2795
Qy 2548 CTCTCAAGCTGTCATGAGATATATGAGAGAGAGATGCTGACACTGATTAATCTC 2607
Db 2796 GTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2855
Qy 2608 AAAGCAAGAGAGATTAAGAACTCTGAGAGAGAGATTAAGAGAGATGCTCCACTCTGT 2667
Db 2856 AAGATTAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2915
Qy 2668 AAACTCTCTCTCCAGCTCCATGACACGCTCTCAACCTCAAGAGAGAGAGAGAG 2727
Db 2916 AAACTCTCTCTCAAGTTCATCAATGCAACCTTATTAATTAATTAATTAATTAATTA 2975
Qy 2728 GTGAGCGTAAAG 2787
Db 2976 GTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3035
Qy 2788 GAGAACTCTCTCTCTCGGCAAGAAATTAAGAACTCAAGAGAGAGAGAGAGAGAG 2847
Db 3036 GAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3095
Qy 2848 TCGTCCGAG 2907
Db 3096 AGTAGTGAAG 3155
Qy 2908 TTCACTAATCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2967
Db 3156 TTTACAAATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3215
Qy 2968 CGTAAGAGCTGAG 3027
Db 3216 AGGAAGAAATTAAG 3275
Qy 3028 CTGTACAAAG 3087
Db 3276 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3335
Qy 3088 AAGTAAAGATGAG 3147
Db 3336 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3395
Qy 3148 AACTCACTGAACATCAAG 3207
Db 3396 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3455
Qy 3208 AAG 3267
Db 3456 AAAG 3515
Qy 3268 CACTCAAG 3327
Db 3516 CATTAATTAAG 3575
Qy 3388 GAG 3387
Db 3576 GAAG 3635
Qy 3388 AAGCTCGAG 3447
Db 3636 AAATTAAG 3695

QY 3448 TCTAGCGGACTGCATCCTGATCGCCGAGCTCAAGGAGTCAATTAGAAACAGAACTAC 3507
 DB 3696 TCAGAGGCTTACATCTTTAAATGTCGATTTAAAGAACTAATAAATAATAT 3755
 QY 3508 ACCGGCAATGCCCAAGCGAAGTAATACAGCTGTAATACGACCTGGAATCTTCAAG 3567
 DB 3756 ACAGGTAATCTCCAAACCGTAATTAATGAGATGTAACAGCTTAAAGATCTTCAAAA 3815
 QY 3568 AATCTCTGCTGAGAGAGATGTCGCCACTGTGCTGTGTAATCTGCTCCGACACA 3627
 DB 3816 AATTTCTCCAGAGAGAAAGATGTTGCAACAGTTCTAAGTGAATGATCCGACACA 3875
 QY 3628 CTGAGACAGTCTCAACCTAAGAAAGCTGCATCTACTCATGTCGAGCGGATCAATACA 3687
 DB 3876 TTGAAACAAGTCAACCAAGAAACAGCATCAATCATGTAGAGCAGAGTCTAACACA 3935
 QY 3688 ATTACCAATCTCAGAAAGCTGACAGATGAGTGCATGATGCTCATGTCCTATCTTC 3747
 DB 3936 ATACCAACATCACAAAATGTCGATGATGAGATGATGATGATGATGATGATGATGAT 3995
 QY 3748 GCGAGAGCGAGAGAGACTACGATGACCTGCGAGGTGTCACCGGTGAGGCTGCTACT 3807
 DB 3996 GGAGATATCCAGAGAGATTTGATGATTTAGGACAGATAGTAACAGAGAGAGAGATCT 4055
 QY 3808 CTTCCGTGATGATTAACATCTCTGCCAAATGAGAGAGAAATGAGAGATGCTATCTG 3867
 DB 4056 ACTTCGTATGATTAACATCTCTGCCAAATGAGAGAGAAATGAGAGATGCTATCTG 4115
 QY 3868 AATCTCTGCGAGCGCTGATGATGCTGCTCAAGAAACAGTGAATGATGATGATGATGAT 3927
 DB 4116 AATCTCTGCGAGCGCTGATGATGCTGCTCAAGAAATGATGATGATGATGATGATGAT 4175
 QY 3928 TTCAATCTCAACGTGAAGACATTTCTGACAGCGCTTTAATGAAGAGAAATTTCAAG 3987
 DB 4176 TTTAATGTTAATGTTAAGATTTTAAATTTCAAGATTTAATGAAGAGAAATTTCAAG 4235
 QY 3988 AACGCTTGAGAGAGCAGCTTGAATTCCTTAATGAAGACCTGACCTCTTAACCTGATGTC 4047
 DB 4236 AATGTTTAAATCAGATTTAATTTCAATTAATGAAGATTTAATGAAGATTTAATGTTGTC 4295
 QY 4048 AAGGACCCATACAAAGTCTCATTAATGAAGAGAGAGAGATTTCTGCTGATGATCAAC 4107
 DB 4296 AATGATTCATTAATTTCTTAATGAAGAGAGAGATTTCTGCTGATGATCAAC 4355
 QY 4108 TATATCAAGGATTCATCGACACCGATATCAATTTGCTAATGATGCTGGGATATTAAC 4167
 DB 4356 TATATTAAGGATTCATCGACATATCAATTTGCTAATGATGCTGGGATATTAAC 4415
 QY 4168 AAGATCTGAGCGAAATATCAAGTCTGACCTGATCTTAATAAAGTATATCAAGAT 4227
 DB 4416 AATATTTATTCGAAATATTAATCAAGTCTGATCTTAATAAAGTATATCAAGAT 4475
 QY 4228 AAGCAAGGAGAGATGAATATCTGCTGATTAATCAATGAGAGAGAGAGATGATCAAG 4287
 DB 4476 AATCAAGGATGAATATGAGAAATCTCTCTTTTAACAAATATGAGAGATGATTAATA 4535
 QY 4288 ACAGTGAAGCAAAATGACCTCTCTGTAATTCACCTGAGCGCAAGTCTCAACTAT 4347
 DB 4536 ACAGTGAATGAATATGATTTATTTGTAATCATTTAGAGAGAAAGTCTCAAAATAT 4595
 QY 4348 ACTTACGAGAGAGATGAGATGAGATTAATAATCAAGAGCTGATCTCAAAACATC 4407
 DB 4596 ACATATGAGAGAAATCAAGTGAAGTAAATTAAGAACTTAATTAATTAATTAATTAAT 4655
 QY 4408 CAAGACAGGCTGAGATTTCAAGAAATTAATTAATTTGAGAAATGAGAGAGCTCT 4467
 DB 4656 CAAGACAAATTTGAGATTTTAAATAAATTAATTAATTTGAGAAATTTGATATTAATCA 4715
 QY 4468 ACCGATTAATACCAACAATCTCTGACCAAGTTTCTGCTCACTGAGAGAGATGATGAG 4527
 DB 4716 ACAGATTAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4775
 QY 4528 AACCTGCGCAAAACAGTGTGAGCAATCTGCTGCA-CGGCAACCTGAGGAGATGCTGAA 4586

DB 4776 AACTTGTAAATCCGTTTATCTTAATTTACTGATGGAACCTTGCAAGGATGTTAAA 4835
 QY 4587 CATCTCCAGCAGCAATGCTGGAAGAAAGAGT--CCCCAGATAGCGCTGTTCCAG 4644
 DB 4836 CATTTCAACACCAATGCTTAAATAAACAATGATCCACAGAGTTCTGATGTTTCA 4695
 QY 4645 CATCTGAGCAGGCGAAGAGTGCATGCTCTCTGTAATCAACAAGAGAGATAG 4704
 DB 4896 CATTTAGATGAAGAGAGATGTAATGTTTAAATTAACAACAAGAGAGATGATA 4955
 QY 4705 TCGTGGAGAGACCAACCTTACTGATGAATGAATAACATGCGGCTGACCGCATGCT 4764
 DB 4956 TGTGTTGAAATTAATCAATCTTACTTGAACGAATTAATGATGATGATGATGATGAT 5015
 QY 4765 AATGACGAG 4824
 DB 5016 AATGATGACGAG 5075
 QY 4825 CCGGACTCTATCCACTCTCTGACAGGATTTTTCCTGACGCTTAATTTCTGGGATC 4884
 DB 5076 CCGATGTTATCCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 5135
 QY 4885 TCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4939
 DB 5136 TCATTTCTTAATTAATCAATGATGATGATGATGATGATGATGATGATGATGATGAT 5190

RESULT 4
 AA080911
 ID AA080911 standard; cDNA; 5181 BP.
 XX
 AC AA080911;
 XX
 DT 24-AUG-1995 (first entry)
 XX
 DE Plasmidium falciparum MSA-1 gene cDNA.
 XX
 XX Plasmidium falciparum MSA-1 gene; recombinant poxvirus;
 KW multicomponent multistage malarial vaccines; immunogens;
 KM malaria diagnosis; ss.
 XX
 OS Plasmidium falciparum (p486195).
 XX
 PN W09428930-A.
 XX
 PD 22-DEC-1994.
 XX
 PF 10-JUN-1994; 94MO-US06652.
 XX
 PR 11-JUN-1993; 930S-0075783.
 PR 09-JUN-1994; 94US-0257073.
 XX
 PA (VIRO-) VIROGENETICS CORP.
 XX
 PI De Taisne C, Paolletti E, Tine JA;
 XX
 DR WPI; 1995-036113/05.
 XX
 PT Recombinant poxvirus contg. Plasmidium DNA in non-essential
 PT region - useful in vaccines against malaria and for prodn. of
 PT Plasmidium immunogens
 XX
 PS Claim 3; Fig 6; 183pp; English.
 XX
 CC AA080911 is the P. falciparum MSA-1 gene cDNA sequence. New
 CC recombinant poxviruses containing either the SERA, ABRA, PfSP70,
 CC AMA-1, PfS25, PfS16, CSP, PfSSP2, USA-1, USA-1 repeats, MSA-1,
 CC MSA-1 (N-terminal p83 or C-terminal gp42) genes, or a combination
 CC of these in non-essential regions of their genomes are claimed.
 CC These poxviruses (pref. with a virulence reducing genomic
 CC deletion or disruption) can be used as vaccines against malaria
 CC and for the prodn. of Plasmidium immunogens. These viruses

CC provide multicomponent, multistage vaccines due to their expression
CC of sporozite, liver stage, blood stage and sexual stage proteins.
XX
SO Sequence 5181 BP; 2287 A; 684 C; 707 G; 1503 T; 0 other;

Query Match 25.5%; Score 1262; DB 16; Length 5181;
Best Local Similarity 56.0%; Pred. No. 1.9e-296;
Matches 2910; Conservative 0; Mismatches 1995; Indels 291; Gaps 16;

```
QY 10 ATGAAATCATTTTCTCTCTGTCATTTTCTGTTTATCATCAATACAGTCGCTG 69
DB 1 ATGAAGATCAATTTCTTTCTATGTCATTTCTTTCTTATTTATTAACAAATGTGTA 60
QY 70 ACCGAGATCCATCAGAGCTGTTAAGAACTGAGAGCTTGGAGCCCTCCCT 129
DB 61 ACACATGAAAGTTATCAAGAACTTCAAAAACAGTAAAGCTTTGAAAGATGCAATTTG 120
QY 130 ACCGATACAGCTGTTCCAGAGAGAGATGCTGTAATGA----- 174
DB 121 ACAGTTATGTTTATTCATTAAGAAAAATGATCTTAATGAAGAAATTAATCTACA 180
QY 175 -----GGGACAGTGGCAGGCGCTTACACACAGCACCCGCT 213
DB 181 AAAGTGCAATGCTCAAAAGTGTAAGTGTAAGTGTAAGTGTAAGTGTAAGTGTA 240
QY 214 TCTAAGGCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 273
DB 241 ACAAGTGTACAGAGTGTACAGAGTGTACAGAGTGTACAGAGTGTACAGAGTGT 300
QY 274 GCTTCGGGCGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 333
DB 301 GGTACAAATGCTACAAAGTGTACAAAGTGTACAAAGTGTACAAAGTGTACAA 360
QY 334 CGAAGAACCATTCATCTGACA----- 355
DB 361 AGTCATATCTCTGTTCAACACTTACCTGTTCAAAATCTTCATCTGTTGCAAGCCT 420
QY 356 -----ACTGAGGATTCGAGCGCAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 408
DB 421 CCAGTGATGAGGATTCAGATGCTAAATCTTACGCTGATTTAAACAGAGATGAGA 480
QY 409 AACATTCCTCCTACTATGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTG 468
DB 481 AATTACTGTTCACTTAATGAAGAACTCAAAATCCGAACTTTGATTTAACTCAATCAT 540
QY 469 ATGCGACACTGTTGATTAATCATGCTCAATATCTGATGAGGTTACCAAGAG 528
DB 541 ATGTTAACTTTGTTGATTAATCATGCTCAATATCTGATGAGGTTACCAAGAG 600
QY 529 ATCAATGAACTCTGATCAAGTGTGATTTCTACTGCTGCTGCTGCTGCTGCTG 588
DB 601 ATTAATGAAATTTATTAATTAATTAATTAATTTTATTTATTAATTAATTAATTA 660
QY 589 GAGCTTGGCGCAATGACTATTTGCAATTCATTTGAAGATGAGAGCCCAAGAG 648
DB 661 GATGATGTGCTAATGATTTATTTGCAATTAATTTCAATTTAAATTCGTCGAATGAA 720
QY 649 TTGAGACGATTTGAAGAGTGTGCTCTCGATATGCAAGCCTCTGCAACATCTAAG 708
DB 721 TTAGACGATCTTAATAAACTGTGTCGATATGAAGAAACCATTTGACAAATTTAAAT 780
QY 709 AATGAGGAAAGATGAGATTTATTTAAAGAAATGAAGAGAGAGAGAGAGAGAG 768
DB 781 AATGAGGAAAGATGAGATTTATTTAAAGAAATGAAGAGAGAGAGAGAGAGAG 840
QY 769 GAGCTGATGAGAGATTCAGAAAGACATAGACAAAGAAATGAAGAGAGAGAGAG 828
DB 841 GAATTAATTTGAAGAGATTTGAAGAAATTTGATCAAAATTAAGAGAGAGAGAGAG 900
QY 829 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 888
DB 901 GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
```

```
QY 889 GAAGCCATTAACCTCATCAGGCTAGTGGAGAGAGCCATAGACACCCCTCAAGAGAGAG 948
DB 961 GAGACATTAATTTATTAAGGCTTTTACAAAAACGATTTAGACACTTTAAAAAATGAA 1020
QY 949 AATATCAAGAACTGCTGAGAGATTAAGAAATTAAGAAATTCCTCGCCAGCAACTCT 1008
DB 1021 AACATTAAGAAATTTCTGATTAAGTAATTAAGTAATTAAGTAATTAAGTAATTA 1080
QY 1009 GGGAAACCCCTTACACGCTGCTGAGCAAGAAAGAGATTAAGAGACAGAGAGAGAG 1068
DB 1081 GAAATTAACCAAAATTAATCTCTGATTAAGAAATTAAGAAATTAAGAAATTAAGAA 1140
QY 1069 ATCAAGAGATGAGCAAAACATTAAGTTCAACATAGATTTCTCTTACTGATCCCTT 1128
DB 1141 ATAAAGAAATTTCCAAACATTAATTAATTAAGATTTGATTTTACTGATCTACTT 1200
QY 1129 GAGCTGAGTACTACTTGAAGAGAGAGATTAAGATTAAGATTAAGATTAAGATTA 1187
DB 1201 GAATTAAGAAATTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1188 -----GACAAAGATTCACCAACCTAATGAATATCCCAATGCTGACATACCT 1239
DB 1261 GATCTACGAAATCTGTTCAAAATTAACCAAAAGTTCTTATCCAAATGATTTGATATCT 1320
QY 1240 CTGCTTATTAACGATATCAACA-----CGCTCAACGAGCTCAATAGCTTCGCT 1290
DB 1321 TTACACCTCAGATATTCATTAATTTATTTAGCTGAGATTAAGATTAAGATTAAG 1380
QY 1291 GACTTGAATTAACCCCTGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350
DB 1381 GATTTAAATGAATCCGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1351 GAGAGAGAGAGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1410
DB 1441 GAAAGAGAGAGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1411 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1470
DB 1501 AATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1471 TATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1530
DB 1561 TATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1531 AACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1590
DB 1621 GTCGATTAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1591 AATACCTTTGATCTATGAGAGATTTCTAAGCATATTTGAGAGAGAGAGAGAGAG 1650
DB 1681 AATAATTTTCTATCTCTATTAATTTCTATTAATTTCTATTAATTTCTATTAATTT 1740
QY 1651 AAGTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1710
DB 1741 TCATATCTTGAAGAGATTTCTTTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1711 AAGATTCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1770
DB 1801 TATATCTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1771 GATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1830
DB 1861 AGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
QY 1831 ATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1890
DB 1918 TCCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
QY 1891 ATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1950
DB 1978 ATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2037
QY 1951 CATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2010
```


Db 2038 CATGTACCAAAATTTATTAACCAAAAATTAACCAAAACATATATTAATGTATTA 2097
QY 2011 AAGAAGAGATGACAAACTGGAAGTGTCTATCCCAAGTCAGAGCCGTGATCAACGA 2070
Db 2098 AAAAAAGAGTATATTAATTAAGAAATTTATACAAAAGTAAGACATGTTAA--GA 2155
QY 2071 GAGAAAGAACATTTAAATGAGAGCAGTGTAGATTAAGTCCGAGCCCTTCACAGAGA 2130
Db 2156 AAGAACAGAGCTGTATATCAAGATTTACACAACTTATAGTGGACAGGAAACAACTG 2215
QY 2131 GAGATACCGGAGAGCTTACCAACCCGGAACAGGCGGTTCAGCTTCGAAGCC 2190
Db 2216 AAGAT-GGGGGGTCTCCACACACATTTATCCCAATCAGGAAACAGAAAGTAAGAA 2274
QY 2191 GATAGGTCGAAGCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2250
Db 2275 GAAACAGAAAGAAAG 2334
QY 2251 GTTCAGAGGCTAAAGCTCAGAGTGCCTACACACAGCTCCTGTGATTAACAGAGAG 2310
Db 2335 CCAAAAGAGTAAAGTGTGTAAGATTTCAATAGAACATTAAGATGACAAATTCACA 2394
QY 2311 AATGTACAGAACTGACATACCTTGAAGAGCTATAGTCTGTGAATACATCTACATC 2370
Db 2395 GCGTTGACAAAACAGTTTATCTAAAGAAATTAAGATGAATTTTAACTAAATCATATA 2454
QY 2371 TGCCAAATATATTCCTGCTCTCAGACAGCTATGAGAGAGAGAGAGAGAGAGAGAG 2430
Db 2455 TGTCATTAATATTTTATAGATCAAACTGATATGACCAAAAATTTATAGAGATAT 2514
QY 2431 AAGATACCAAG 2490
Db 2515 AATCTTACCCAGAAAG 2574
QY 2491 AATATCAG 2550
Db 2575 AATATTCAG 2634
QY 2551 TCTCAACTGTTCATGAGATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2610
Db 2635 CAACATCTCTTTTGAATATATCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2694
QY 2611 GACAG 2657
Db 2695 GAGGAG 2754
QY 2658 -----CACCTGTGTTAAACCTGCTCTTC 2681
Db 2755 ACATTCAGTCTGGAATATACACCGTAATATCTGCTCAATCCGCAACTCAGATATTC 2814
QY 2682 CAGCTCATGCAACCACTGTCTCTCACACTCAAGACA----- 2720
Db 2815 CAACCAACCAACCAATCAATGATCCTCTACCAATACCAAAATGGTATGATCATCT 2874
QY 2721 -----GCCGGAAGTGAAGCGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2775
Db 2875 GGTCTGCTGTATGAG 2934
QY 2776 CTGAAGCTGTTGAGAACTCTGTCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2832
Db 2935 TTGAAGAGATTTTGTAGTCTCTTAATCTTGGAATTAAGAGAGAGAGAGAGAGAGAGAG 2994
QY 2833 CTATTTGAG 2892
Db 2995 ACCATTCTTCACACAG 3054
QY 2893 TTCTATAG 2952
Db 3055 TATTTTAATGATATATCAACCAATTCGTAAATCTTAATTCAAAGAGAGAGAGAGAGAG 3114
QY 2953 AAGGATGAATCTAAGGTGAAG 3012
Db 3115 ACC---GAAACACAAAAAATGATTAATGATTAATTAATTAATTAATTAATTAATTA 3171
QY 3013 CAACGTAGCTGTGACCTGTATACACAGATCAAACTGAACTGAGAGAGAGAGAGAGAGAG 3072
Db 3172 CAGTTTATCATTTGATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3231
QY 3073 AAG 3132
Db 3232 AAAAAAG 3291
QY 3133 CTGAAG 3192
Db 3292 TTGAATCAAAATTTGAATTTCAATTTAATCAATTAATTAATTAATTAATTAATTAAT 3351
QY 3193 TTCTTACAG 3252
Db 3352 TTCTTACAG 3411
QY 3253 AAGATTTCTCTCAACAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3312
Db 3412 AAAAAATTTATTTGAACATTTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3471
QY 3313 CTGAAG 3372
Db 3472 TTAAACCTTTAAG 3531
QY 3373 TTCAAGCTCTGTCTAAGCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3432
Db 3532 TTATAGATTTATTAAGATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3591
QY 3433 AAGCTAG 3492
Db 3592 AATTTCTCTTTCTATCAAG 3651
QY 3493 AAG 3552
Db 3652 AAAAAATTAATTAATTAAG 3711
QY 3553 CTGAAGATCTTCAAG 3596
Db 3712 TTAAATCTTACCAAAATTTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3771
QY 3597 -----CACGTGAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3641
Db 3772 CAACAGAGATTAATCTCATCTCATTTCTGTAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 3831
QY 3642 ACCTAAGAA-----GCTGATCTATCTATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3693
Db 3832 AAG 3891
QY 3694 ACATCTCAG 3753
Db 3892 CAATTTCAAAATTTATGAG 3951
QY 3754 AGGAG 3813
Db 3952 TCCGAGAGATATAG 4008
QY 3814 GTGATGATTAATCTTCTGTCACAAATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3873
Db 4009 ACATGATTAATATCTCTCTCAG 4068
QY 3874 CTGCGAG 3933
Db 4069 TTAGCTGAGATATATGAG 4128
QY 3934 GTCAAGCTGAG 3993
Db 4129 TTAAATTTGAGAGATATCTTAATTTCAAGCTTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4188
QY 3994 TTGAG 4053
Db 4189 TTGAATCTGATTTAATGCAATTTAAG 4248

XX Davidson EA, Yang S;
 XX
 XX WPI: 1997-393372/36.
 DR
 XX Malaria vaccine - comprises expression vector expressing fragment of
 PT merozoite surface antigen
 XX
 PS Claim 33; Fig 4; 75pp; English.
 CC The present sequence encodes a chimeric protein that contains a
 CC human anchor sequence, and a merozoite surface antigen-1 (MSA-1) carboxy
 CC terminal peptide. The C-terminal fragment of MSA-1 provides a more
 CC specific response than the complete MSA-1, and attachment of the
 CC anchor sequence improve the immunogenicity of the protein better than
 CC the use of an adjuvant. The chimeric protein, and expression vectors
 CC (analogues that express MSA-1 without either signal or anchor peptides),
 CC particularly in the form of recombinant vaccinia virus, are used in
 CC vaccines to prevent or treat malaria caused by Plasmodium falciparum. The
 CC vaccinia vector expresses the antigen fragment for many days, or even
 CC years, generating a long-lasting immune response (humoral and/or
 CC cell-mediated) against the merozoite form of the parasite, in humans or
 CC other animals.
 CC
 SQ Sequence 1897 bp; 815 A; 230 C; 274 G; 578 T; 0 other;
 Query Match 20.0%; Score 989; DB 18; Length 1897;
 Best Local Similarity 72.3%; Pred. No. 3,2e-230;
 Matches 1297; Conservative 0; Mismatches 495; Indels 1; Gaps 1;

QY 3138 AAGCAAACTCACTCACTGACATCCGAACACGTAAGTCTGAGTCTT 3197
 DB 105 AATGATGATGATCTCACTTAATACCCAAAGCATGATTAACAACCTTTCTGTTCTT 164
 QY 3138 CAACAAGAAAGAAAGCCGAGATCCGAGACAGAAACACTCTGGAGACACCAAGAT 3257
 DB 165 TAACAAAAAAGAAAGCTGAAGTACGAAAACTGAAAAACATTAGAAAAACAAAAAT 224
 QY 3258 TCTTCTCAACACTCAAAAGGCTGCTCAAGTATATATATGCGAGTCTTCCCTGTA 3317
 DB 225 ATTATTTGAACACTTAAAGGACTTGTATATATATATATATATGATGATCTTCATTA 284
 QY 3318 GACTCTCCGAGAGAGATCCAGACCGAGATTAAGTACGACCTTCGAGAGACTTCA 3377
 DB 285 AACTTTAGTGAAGATCAATTCAAACAGAAAGATATATATGCAATTTGAAGAACTTAA 344
 QY 3378 GGTCTGCTTAAGCTGGAAGCAAGCTGAAGCAACTGAACTGGAGAGAAAGAGCT 3437
 DB 345 AGTATTAAGTAAATTAAGAAAGAAATTAAGATTAATTAATTAAGAAAGAAATTT 404
 QY 3438 CAGCTACCTCTAGCGAGCTGCATCACTGATCCCGAGCTCAAGAAAGTATTAAGA 3497
 DB 405 ATCATACCTTATCAAGTGGATTAATCATATTAATTCCTGAATTAAGAAAGTATTA 464
 QY 3498 CAAGAATAACACCGCAATAGCCCAAGCAGAGATATATAGAGTGAATTAAGCACTGA 3557
 DB 465 TAAATATTAATACAGTAAATTCCTCAAGTAAATATACGATGTTAAACAATGCAATTA 524
 QY 3558 ATCTTACAGAAAGTCTGCTGGAAGAAACAGATGTCGCCACTGTGGTGTCTGAATCT 3617
 DB 525 ATCTTACAAAAAATTTCTCCAGAAAGAAACAGATGTTGCAACAGTGTAAAGAAAGT 584
 QY 3618 CTCGACACACTGAGAGCTGCATCACTGAAGCGCTGATCTACTCATGCTGGAGCGCA 3677
 DB 585 ATCCGACACATTTAGAACATGCAACCAAGAAACAGATCACTCATGATGAGGACAA 644
 QY 3678 GTCCATATATATACACATCTTCAGAACCTCGACATAGTGTGATGATGATGATGAT 3737
 DB 645 GTCTAACACATTAACACATCAACAATATGATGATGATGATGATGATGATGATGAT 704
 QY 3738 GCTATCTTGGCGAGAGCGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 3797
 DB 705 ACCTATATTTGGAGATCCGAAAGAAAGATATGATGATTTAGGACAAAGTATGATGAT 764

QY 3798 GGCTGCACTCTCTCCGATGATTAACATCTGTCCAAATCGAAGCAAGTAAAGT 3857
 DB 765 AGCAGATACCTCTCCGATGATTAACATCTGTCCAAATCGAAGCAAGTAAAGT 824
 QY 3858 GCTTATCTGAACCTCTGCGAGCGCTATATAGTCTCTCAAGAAACAGCTGGAGATTA 3917
 DB 825 TTTATATTTAAACCTTTAGCAGGTGTTATGAGTTTAAACCAATTTGAAAAATTA 884
 QY 3918 CGTATGACCTTCAATGTCACAGTGAAGACATCTGCAACGCCCTTATTAAGAGAGA 3977
 DB 885 CTTATGACATTTATGTTAAATGTTAAGGATTTTAAATTCACGATTTAAATTAACG 944
 QY 3978 AATTTCAAGACGCTTGGAGAGAGCTGATGCTCCATTAATTAAGACCTGACCTCTTA 4037
 DB 945 AATTTCAAAAAATGTTTATGAAATCAAGATTTATCTCATTAATTAAGATTTAACA 1004
 QY 4038 CTACGTTGCAAGAGCCATACAGTCTCTCAATTAAGAGAGATTAATTTCTGTC 4097
 DB 1005 TTATGTTGCAAGATCCATTAATTTCTTATTAAGAAAAAGAGATTAATTTCTTA 1064
 QY 4098 TAGTTACAACTTATATCAAGAGCTGCACACGATATCAATTTGCTAATGATGCT 4157
 DB 1065 CAGTTATATTTATTTAAGATTTCAATAGATAGATTAATTTGCAAAATGATGCT 1124
 QY 4158 GGGTATTTACAGATCCGAGGAAATTAACAGTCTGACCTGCTATTAATTAAGATA 4217
 DB 1125 TGGATTTATTAATTAATTTATTCGAAAAATTAATTAAGATTTAATTAATTAATTA 1184
 QY 4218 TATCAACGATTAAGCAAGGCGAATGAATGAATATATGCTCTTCAATTAACATGA 4277
 DB 1185 TATCAACGCAACAAAGTGAAATGAGAAATACCTTCCCTTTTAAACAATTTGAGAG 1244
 QY 4278 CCTGTAACAGAGTGAAGCAAGCAAAATGACCTCTTCTGATTTACCTGGAGGCCAGT 4337
 DB 1245 CTTATATTAACAGTAAATGATTAATTAATTAATTTGATTAATTTAAGAACAAAGT 1304
 QY 4338 CCTCACTACTTATTAAGCAAGATGGAAGTGAAGTAAATTAACAGGAGCTGAATCCT 4397
 DB 1305 TCTAATTTATATATTTGAGAAATCAACGTTAGAGTTAAATTAAGAACTTAATTA 1364
 QY 4398 CAAACAAATCCAAAGCAAGCTGCGAGATTTCAAGAAAAATTAACAATTTCTGGA 4457
 DB 1365 AAAAAAATTCAAAGCAAAATGCGAGATTTTAAAAAATTAACAATTTCTGGAATG 1424
 QY 4458 AGACTGCTTACCGATTTATTAACCAACAATCTCTGACCAAGTTTCTGCTCAGTGC 4517
 DB 1425 TGATTTATCAACAGATTTATTAACCAATTAATTAATTTGACAAAGTTCTTAAG 1484
 QY 4518 GGTGTTGAAACCTCCGCAAAACAGTCTGAGCAATGCTCGAGGCAACCTGACGG 4577
 DB 1485 GGTGTTGAAACCTGCTGAAACAGTCTTATTAATTTACTGTTGGAACCTTCAAG 1544
 QY 4578 CATGCTGAACATCTCCAGCACCAATGCGTGAAGAAAGAGTCCGCCAATATGCGGT 4637
 DB 1545 TATGTTAAACATTTCAACAACCAATGCGTAAAAAAACAATGTCACAAAATTTCTG 1604
 QY 4638 TTTCAAGGATCTGAGCAGCGCGGAAGAGTGAAGTCTCTGAACTACAAACAAGAG 4697
 DB 1605 TTTCAAGACATTTAGATGAAGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1664
 QY 4698 AGATTAAGCGGTGAAGAACCCCAACCTACCTGCAANTGAAGAAACANTGGGGTGA 4757
 DB 1665 TGATTAATGTTGTTGAATATCCAAATCTCTGTTATTAAGAAATTAATGTTGAT 1724
 QY 4758 CGATGCTAAATGCAACGAGGAAGACAGCGGCTTAACGGAAGAAATCAATGCGAG 4817
 DB 1725 AGATGCAAAATGATCGAAGAGATTTCAAGGTAGCAAGGAAAGAAATCAATGCGAG 1784
 QY 4818 TACTAAGCCGACCTCTATCCACTCTTGAGCGGAGATTTTCTGCTCAGCTCAATTTCT 4877
 DB 1785 TACTAAGCCTGATCTTATCCACTTCTGATGATTTTCTGCAAGTCTCTCAATTTCT 1844


```
OY 4278 CCTGTACAGACGAGCAGCAACCAATCGACCTCTGTAATTCACCTGGAGGCCAAGCT 4337
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1299 CTTATATTAACAGATTATGATTAATTAATTTATTTGTAATTCATTGAAACCAAGT 1358
OY 4338 CCTCACTACTACTTACGAGAGCAATGTGCAATTTAAATTCAGAGCTGAATACCT 4397
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1359 TCTAATATATATGATGAAATTCAGAGTAGAAGTTAAATTAAGAACTTAATTAAGT 1418
OY 4398 CAACAATTCAGAGACAGCTGGCAGATTTCAGAAATAATACATTTGTCGGAATTC 4457
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1419 AAAAACAATTCAGAGCAATTTGGCAGATTTTAAAAAATTAACATTTGTCGGAATTC 1478
OY 4458 AGACCTGTACGATTTATACCAACAATCTCCAGCAATCTTCCAGCTGGCAAT 4517
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1479 TGTATTTATACAGATTTATACCATTAATTAATTAATTCCTTAGTACAGTAT 1538
OY 4518 GGTGTTCGAAAACCTGCGCAAAACAGTGTGAGCAATCTGTCGAGCGCAACCTGAGGG 4577
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1539 GGTTTTGAATAATCTGTAAACCGTTTATCTAATTTACTGATGAGAACTTGCAGAG 1598
OY 4578 CATGCGAATCTCCACAGCAGCAATGCTGGAAGAAAGTCCGCCAATAGCGGTG 4637
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1599 TATGTTAAATCTTCAACACCAATGCGTAAAAAACAATGTCCAAAAATTTCTGGATG 1658
OY 4638 TTTCAGGCACTTGACAGCGGCAAGAGTGCAAGTCTCTGCAATACAAACAAGAG 4697
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1659 TTTCAGACATTTAGATGAAGAAGAAAGATTAATTTTATTAATTAACAACAAGAG 1718
OY 4698 AGATTAAGCTGTGAGACCCAAACCTTACTGCAATGAAGAAACAATGGCGGTGACGC 4757
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1719 TATATAATGTGTGAATAATCAAAATCTTACTGTAACGAAAAATTAATGATGATGC 1778
OY 4758 CGATGCTAATGACGAGAGAGACAGCGGCTTAACGAAAAAATTAACATGCGAGTG 4817
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1779 AGATGCAATATGACGAGAGAGATTCAGGTACCAACGAGAAATAATCATGTGAATG 1838
OY 4818 TACTAAGCCGAGCTCTATTCACCTCTTCGACGGGATTTTGGCTCCAGCTCAAT 4873
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1839 TACTAAGCTGATTTTATTCACCTTTTCGATGATTTTCTGCAAGTTCCGTCACT 1894

RESULT 8
AA56008
ID AA56008 standard; cDNA; 1065 BP.
XX
AC AA56008;
XX
DT 31-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-1-42 modified cDNA.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KM protein engineering; codon usage; transgenic animal; ss.
XX
OS Plasmodium falciparum.
XX
PN MO9920774-f2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
XX
PR 20-OCT-1997; 97US-0062592.
XX
PA (GEN2 ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-28813/24.
XX
P-PSDB: AAY09372.
```

```
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Claim 1: Fig 1: 35pp; English.
XX
CC This novel, modified nucleic acid encodes the 42 kDa C-terminal
CC portion (see AAY09372) of malaria merozoite surface protein MSP-1
CC (MSP-1-42), an important target for the development of a vaccine
CC against Plasmodium falciparum. The nucleic acid sequence has been
CC modified compared to the native sequence of MSP-1-42 (see AA56009)
CC such that 306 nucleotide positions have been replaced to lower the
CC AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
CC motifs while maintaining the same protein amino acid sequence.
CC These alterations allow MSP-1-42 to be expressed in mammalian cell
CC culture and in transgenic mice. Native MSP-1-12 is known to be
CC difficult to express in cell culture systems, mammalian cell
CC culture systems or in transgenic animals. The invention allows
CC expression of MSP-1-42 protein in the milk of transgenic animals,
CC and also provides a DNA vaccine comprising a vector containing the
CC altered MSP-1-42 sequence.
XX
SQ Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;
XX
Query Match 15.1%; Score 744.2; DB 20; Length 1065;
Best Local Similarity 81.3%; Pred. No. 9, 3e-171;
Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;
OY 3799 GCGTGTACATCTCCCTGATGTAATTCATCTGCAAAATCGAAGCAATGCAAGT 3858
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 GCCGTACATCCCTCCCTCATGTAATCTGTAATCTGCAAGTCAAGACGATGCGAGTG 60
OY 3859 CTTCTATCTAATAACCTTGGCAGCGCTGTATAGTCTCTCAAGAAACAGCTGGAATAAC 3918
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 CTGTACTGTGAAGCGCGGTGGAGGGGTCTACCGAGCTGAAGAGAGCTGGAGAACAC 120
OY 3919 GTGATACCTTCAATGTCAACGTAAGAGCATTTCTAAGACCGCTTATTAAGAGAA 3978
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 121 GTGATACCTTCAACGTGAACGTAAGAGATATCTCAACAGCGGTTCAACAGCGGAG 180
OY 3979 AATTTCAGAGCTCTGGAGAGCGATGATTCCTTAAAGACCTGACCTCTCTAAC 4038
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 AACTTCAGAGCTGTGGAGAGCGATGTATCCCTTCAAGAGATGTGACAGCACACAC 240
OY 4039 TACGTGTCAAGAGACCATACAGATTCTCAATTAAGAGAGAGGATTAATTTCTGTCT 4098
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 241 TACGTGTCAAGAGATCCCTACAGATTCTCAAGAGAGAGAGAGATTAATTTCTGTGAGC 300
OY 4099 AGTTACACTATATCAAGAGATCTCAAGACACCGATATCAATTTGCTTAATGATGCTG 4158
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 301 AGTTACACTATATCAAGAGATCTCAAGATGATGATATCAATTTGCTTAATGATGCTG 360
OY 4159 GGGTATTCAGAGTCTGGAGAGCGAATAATCAAGTGTGACCTTGTCTTAAAGATAT 4218
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 361 GGATTAATTCAGAGTCTGGAGAGCGAATAATCAAGTGTGACCTTGTCTTAAAGATAT 420
OY 4219 ATCAAGATTAAGAGAGGAGGAATGAATAATATCTCCCTCTGATTAATCAAGAAC 4278
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 ATCAAGATTAAGAGAGGAGGAATGAATAATATCTCCCTCTGATTAATCAAGAAC 480
OY 4279 CTGTACAGAGTGAAGAGCAATATGACCTTGTGATTAATCAAGAGGAGGAGTTC 4338
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 481 CTGTACAGAGTGAAGAGCAATATGATGATGATGATGATGATGATGATGATGATGAT 540
OY 4339 CTCAACTATATTAAGAGAGCAATGGAAGTGAATTAATCAAGAGGAGGAGTGAATCTC 4398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 541 CTCAACTATATTAAGAGAGCAATGGAAGTGAATTAATCAAGAGGAGGAGTGAATCTC 600
OY 4399 AAAACAATCAAGAGAGGAGGAGTTCAGAAAAATTAACATTTCTCGAATTTGA 4458
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 601 AAGACCATCAAGATTAAGTGGCGGATTCAGAAAGAACACAACTTCGCGGAGTGGC 660
OY 4459 GACCTGTACAGATTAATCAACAACAACTTCGCGGAGTTCGCGGAGTTCGCGGAGT 4518
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 661 GATCTGAGCAGCAGATTAACAACAACAACTTCGCGGAGTTCGCGGAGTTCGCGGAGT 720
```


Qy	4519	GTGTTGAAAACTCTGCCAAACAGTGTCTGAGCAATCTCTTGACGGCACCCTGCAGGGC	4578
Db	721	GTTCTTCAAAACCTGGCCCAAGACCCTGTCTTGAGCAACCTCTGGATGGGACCTGCAGGGG	780
Qy	4579	ATGTGGAACATCTCCAGACCAATGCGTAAGCAAAACAGTGGCCCCAGATTAGCGGCTGT	4638
Db	781	ATGTGGAACATCAACCCAGCACACAGTGTGTGAACAAACAATGTGCCAGACACGGGTGT	840
Qy	4639	TTCAGCATCTGCAGCAGGCCAGGAAGTCAAGTGTCTCTGAATCTACAAACAAGAA	4698
Db	841	TTCAGACACCTGGATGTAGAGAGAGAGAGTGAATGTCTCTCTAATAACAGAGAAAGT	900
Qy	4699	GATAAGTGGCTGTGGAACCCCAACCCCTACTCTGCATGAAGAAACAAATGGCGGTGTACGCC	4758
Db	901	GATAAGTGTGTGGAAAACCCCAATCTCTACTTGTAAAGAGAACAAATGTGGATGTATGCC	960
Qy	4759	GATGCTAAATGTCAACCCAGGAAGACAGCGGCTTAACGGAAGAAATCATCATGCGATGT	4818
Db	961	GATGCAAGTGTACCGAGGAGGATTCAGAGGAGCAACGGGAAGAAATCATCATCTGTGATGT	1020
Qy	4819	ACTAAGCCGCACTCTATCCACACTCTCGAGCGGATTTTTTG	4859
Db	1021	ACCAAGCCTGATTTTATTCACAGTTCGATGAGGTATCTTTTG	1061

RESULT 9	
AAx25586	
ID	AAx25586 standard; cDNA; 1065 BP.
XX	
AC	AAx25586;
XX	
DT	02-AUG-1999 (first entry)
XX	
DE	Merozoite surface protein MSP-1-42 modified cDNA.
XX	
KW	MSP-1; merozoite surface protein; malaria; vaccine;
KW	protein engineering; protein expression; codon usage;
KW	transgenic animal; ss.
XX	
OS	Plasmodium falciparum.
OS	Synthetic.
XX	
PN	W09920766-A2.
XX	
PD	29-APR-1999.
XX	
PE	20-OCT-1998; 98WO-US22225.
XX	
PR	15-MAY-1998; 98US-0085649.
PR	20-OCT-1997; 97US-0062592.
XX	
PA	(GEN2) GENZYME TRANSGENICS CORP.
XX	
PI	Chen LH, Meade H;
XX	
DR	WPI: 1999-302742/25.
DR	P-PSDB; AAY05832.
XX	
PT	New modified recombinant nucleic acid sequences useful for producing
PT	malarial DNA vaccine
XX	
PS	Claim 12; Fig 1; 43pp; English.
XX	
CC	This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC	part (see AAU05832) of malaria merozoite surface protein MSP-1
CC	(MSP-1-42), an important target for the development of a vaccine
CC	against Plasmodium falciparum. The nucleic acid sequence has been
CC	modified compared to the native sequence of MSP-1-42 (see AAx25587)
CC	such that 306 nucleotide positions have been replaced to lower the
CC	AT content (from 76 to 49.7%) and to eliminate 10 mRNA instability
CC	motifs while maintaining the same protein amino acid sequence.
CC	These alterations allow MSP-1-42 to be expressed in mammalian cell

CC culture and in transgenic mice. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, mammalian cell culture systems
CC or in transgenic animals. The preferred difficult protein
CC candidates for expression are those derived from lower organisms
CC such as parasites, bacteria and viruses that have DNA coding
CC sequences of high AT content or which have mRNA instability motifs
CC or rare codons relative to the recombinant expression system to be
CC used. The invention allows expression of MSP-1 protein in the milk
CC of transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.

Sequence 1065 BP; 324 A; 254 C; 280 G; 207 T; 0 other;

Query Match	15:18;	Score 744.2;	DB 20;	Length 1065;
-------------	--------	--------------	--------	--------------

```

best local similarity  0.15; 100: 0.15 17.1;
Matches 863; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

```

QY 3799 GCTGCACTCCTTCGATGATAACATTCTGTCCAAATCGAGACGATACGAAGTG 3858

Db 1 GCCGCTACTCCCTCGTCATCGATAACATCCTGTCCAGATCGAGACGAGTACGAGGTG 60

QY 3859 CTCTATCTGAACCTCTGGCAGGCGTCTATAGGTCCTCTCAAGAAACAGCTGGAGATAAC 3918

Db 61 CTGTACCTGAAGCCGCTGGCAGGGGTTACCGGAGCCTGAAGAAGCAGCTGGAGAACAAC 120

3919 GTGATGACCTTCATGTCAACGGAAGACATTCTGAACAGCCGCTTTAATAAGAGAGAA 3978

Db 121 GTGATGACCTTCAACGTGAACGATATCCTGAACAGCCGGTTCAACAAGCGGAG 180

3979 AATTCAAGACGCTTGGAGAGCGACTTGATTCCCTATAAGACCTGACCTCCTAAC 4038

Db 181 AACTTCAGAACGTGCTGGAGAGCGATCTGATCCCCCTACAGGATCTGACCAGCAGCAAC 240

QY 4039 TACGTGTCAAGGACCATACAGTTCCTCAATAAGAGAAGAGGATAATTCTGCT 4098

Db 241 TACGTGTC AAGGATCCCTAC AAGTTCCTGAACAAGGAGAGAGATTAAGTTCCTGAGC 300

QY 4099 AGTTACACTATATCAAGGACTCCATCGACACC GATATCAATTTCGCTAATGATGTGCTG 4158

Db 301 AGTTACACTACATCAAGGATAGCATTGATACCGATATCAACTTCGCCAACGATGTCCTG 360

QY 4159 GGGTATTACAGATCCTGAGCGAAAAATACAAGTCTGACCTTGACTCTATTAAAAAGTAT 4218

Db 361 GGATACTACAAGATCCTGTCCGAGAAGTACAAGAGCGATCTGGATTCAATCAAGAAGTAC 420

QY 4219 ATCAAGATAAGCAAGCGAGAATGAAAAATATCTGCCCTTCCTGAATACATCGAAACC 4278

Db 421 ATCAACGATAAGCAGGAGAGAACGAGAAGTACCTGCCCTTCCTGAACCAACATCGAGACC 480

QY 4279 CTGTACAAGCAGTGAACGACAAATCGACCTCTTCGTAATTCACCTGGAGGCCAAGGTC 4338

Db 481 CTGTACAGACCGTCAACGATAAGATTGATCTGTTGCGTATCCACCTGGAGGCCAAGGTC 540

QY 4339 CTCACCTATACCTTACGAGAGAGCAATGTGGAAGTTAAATCAAGGAGCTGAACCTACCTC 43398

Db 541 CTGAACTACACATATGAGAAGAGCAACGTGGAGGTC AAGATCAAGGAGCTGAATTACCTG 600

4399 AAAACAATCCAAGACAAGCTGGCAGATTTCAGAATAACAATTTCGTCGGAATTGCA 4458

Db 601 AAGACCATCCAGGATAAGCTGGCCGATTTCAGAAGAACACAACTTCGTCGGGATCGCC 660

4459 GACCTGCTACCGATTATAACCAACAATCTCCTGACCAAGTTTCTGTCCACTGGCATG 4518

Db 661 GATCTGAGCACC GATTACAAACCACAACTGCTGACCAAGTTCCTGAGCACC GGTA TG 720

4519 GTGTTGAAACCTCGCCAAACAGTGTGTCGACCAATCTGCTCGACGGCAACCTGCAGGC 4578

Db 721 GTCTTGAACCTGGCCAGACCGTCTGAGCAACCTGCTGGATGGGAACCTGCAGGG 780

QY 4579 ATGCTGAACATCTCCAGCACC AATGCGTGAAGAACAGTGC CCCCAG AATAGCGCTGT 4638


```

Db 781 ATGTCGAACATCAGCCACCAGCTGTGTGTAAGAACAGATGTCCCAAGAACAGCGGTGT 840
Oy 4639 TTCAGGCATCTGGACGCGGAGAAAGTGCAGTCTCTCTTAATACAAACAAAGAGA 4698
Db 841 TTCACACCTGTGATGAGAGAGAGTGAAGTCTGTCTGTAATACATACAGAGAGGT 900
Oy 4699 GATAAGTCGTGAGAACCCCAACCTCTACCTGCAATGAACAGCGGTGTGAGCC 4758
Db 901 GATAGTGTGTGAGAAACCCCAATCTCTTAACGAGAACATGTGTGATGTGCC 960
Oy 4739 GATGCTAAATGCACCGAGAGACAGCGCTCTTAACGGAAGAAATCATGCGATGT 4818
Db 961 GATGCAAGTGTACCGAGAGAGATCTCAGGAGAACGGAAGAAATCATCTGTGATGT 1020
Oy 4819 ACTAGCCGCACTCTATCCACTCTTCGACGCGGATTTTGTG 4859
Db 1021 ACCAAGCCTGATCTTATTCACCTGTTCATGTGTCTCTG 1061

RESULT 10
AA56021
ID AA56021 standard; cDNA: 1140 BP.
AC AA56021:
XX 31-AUG-1999 (first entry)
DT
XX Merozoite surface protein MSP-1-42 modified cDNA.
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; mutant; ss.
XX Plasmodium falciparum.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH CDS 1..1131
FT /*tag= a
FT sig_peptide 1..45
FT /*tag= b
FT mat_peptide 46..1128
FT /*tag= c
XX
XX MO9920774-A2.
XX
XX 29-Apr-1999.
XX
XX 20-OCT-1998; 98WO-US222226.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX WPI: 1999-288313/24.
XX P-PSDB; AAI09374.
XX
XX Modified malarial protein for use in anti-malarial vaccines
XX
XX Example; Fig 11; 35pp; English.
XX
XX This novel, modified nucleic acid encodes a 42 kDa C-terminal
XX pat (see AAI09374) of malaria merozoite surface protein MSP-1
XX (MSP-1-42), an important target for the development of a vaccine
XX against Plasmodium falciparum. The nucleic acid sequence has been
XX modified compared to the native sequence of MSP-1-42 (see AA56009)
XX such that the AT content has been reduced and 10 mRNA instability
XX motifs eliminated while maintaining the protein amino acid sequence.
XX In addition, a sequence encoding a 15-amino acid beta-casein signal
XX peptide has been added to the 5' end of the sequence, and N262Q

```

```

CC and N181Q mutations have been introduced to eliminate
CC N-glycosylation sites. These alterations allow MSP-1-42 to be
CC expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC invention also provides a DNA vaccine comprising a vector containing
CC an altered MSP-1-42 sequence.
XX
SQ Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;
Query Match 15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity 80.8%; Pred. No. 3.3e-170;
Matches 865; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
Oy 3791 CCGGTGAGCGCTGCATCTCTCCGTGATGATTAATCTGTCCAAATCGAAGCAAT 3850
Db 38 CCATTGCAAGCGCTCATCTCCCTCCGTCATCATATCATCTGTCCAAAGATCGAAGCAAT 97
Oy 3851 ACGAAGTGTCTATCTGAAACCTCTGGCAGGCGCTGTATAGTCTCTCAAGAAACGCTGG 3910
Db 98 ACGAGTGTCTTACTGAAAGCCCTGGCAGAGTCTACAGAGCCTGAAAGCAGCTGG 157
Oy 3911 AGAATACGCTGATGACCTTCATATGTCACAGTGAAGACATTTGTGAACGCCCTTAAAT 3970
Db 158 AGAACACGCTGATGACCTTCATATGTCACAGTGAAGACATTTGTGAACGCCAGTTCAC 217
Oy 3971 AGAGAGAAATTTCAAGAACTCTTGGAGAGCGACTGTATCTCTTAAAGACCTGACT 4030
Db 218 AGAGGAGAACTTCAAGAAAGCTGTGAGAGCGATCTATCCCTACAAAGATCTGACCA 277
Oy 4031 CCTTAATACGTTGTCAAGAACCCATACAGTTCCTCATATTAAGAAAGAGATTAAT 4090
Db 278 GCAGCAACTACGCTGTCAAGAAAGATCCCTACAACTCTCGAACAAGAGAAAGATTAAGT 337
Oy 4091 TTCTGTAGTATACAACTATATACAGAGACTCCATGACACCGATATTCATTTCCGTAATG 4150
Db 338 TCTTAGACGATTAATATATATATAGATAGCATTTGACACGATATTCATCTTGCCACAG 397
Oy 4151 ATGTCGTGGGTATTACAAAGATCTGTAGGAAATATACAACTGTACCTTGACTATTTA 4210
Db 398 ATGTCCTGGGATATACAAAGATCTGTCCGAAATATACAAAGCGATCTGTATAGATCA 457
Oy 4211 AAAATATATATACAGTAAAGCAGGCGAGATGAAATATATGTGCCCTTCTGTAATACA 4270
Db 458 AGAAGTACATCAACGATTAAGCAGGAGAGAAAGAAAGATCTGCCCTTCGAACACACA 517
Oy 4271 TCGAAACCCGTACAAAGCAGAGAACGCAAAATGACCTCTGTAATTCACCTGGAGG 4330
Db 518 TCGAAGACCCGTACAAAGACGTCACAGATTAAGATTTGATCTGTGATCCACCTGGAGG 577
Oy 4331 CCAAGTCTCTCACTATCTTACGAGAGAGCAATGTGAAATTAATCAAGAGCTGA 4390
Db 578 CCAAGTCTCTGATACATATATGAAAGAGCAAGCTGAGGTCAAGATCAAGAGCTGA 637
Oy 4391 ACTACCTCAAAACAATCCAAAGCAAGCTGGCAGATTTCAAGAAATTAACAATTCGTGC 4450
Db 638 ATTACCTGAAGACCATCAGATTAAGCTGGCCGATTTCAAGAAACAACAATTCGTGC 697
Oy 4451 GAATGCAAGACCTGTCTACCCGATTATACCAACAACAATTCCTGACCAAGTTTGTGTCA 4510
Db 698 GAATGCGCATGTGAGACCCGATTAACAACAACAATTCCTGACCAAGTTTGTGTCA 757
Oy 4511 CTGGCATGGTGTTCGAAAAACCTCGCCAAACAGTCTGAGCAATCTGCTCGAGCAAC 4570
Db 758 CCGGAATGGTCTTCGAAAAACCTCGGCCAAGACCGTCTCTGAGCAACTGCTGTGGAAC 817
Oy 4571 TCGAGGCATGCTGAATCTTCCAGACCAATTCGCTGAAGAAAGATGCCCCAGAAATA 4630
Db 818 TCGAGGCATGCTGAATCTTCCAGACCAATTCGCTGAAGAAAGATGCCCCAGAAACA 877
Oy 4631 GCGGCTGTTTCAGGATCTGAGACGAGCGCAAGAGTGAAGTCTGTAATCTCAAAAC 4690
Db 878 GCGGATGCTTCAAGACCTGATGAGAGGAGAGATGCAAGTGTCTGTAATCTCAAGC 937
Oy 4691 AAGAGAGAGATTAAGTCTGAGAACCCAAACCTTACTGTCAATGAAACCAATGCGGGT 4750

```



```

Db      938 AGGAAGAGATAGTGTGGAAGAACCCCAATCTACTTACGAGAACATGAGAGAT 997
Oy      4751 GTGACCCCATGCTAAATGACACGAGAACAGCGGCTTAACGAAAGAAATACAT 4810
Db      998 GCGATCCCGATGCGCAAGTGACGAGAGATTTCAGAGAACGAAAGAAATACACT 1057
Oy      4811 GCGAGTGTACTAAGCCGACCTCTATCATCTCTTCGACGAGGATTTTTCG 4860
Db      1058 GCGAGTGTACCAAGCCGATTTCTATCATCTGTTCAGATGATTTTCTG 1107

RESULT 11
AAK25593
ID      AAK25593 standard; cDNA; 1140 BP.
XX
AC      AAK25593;
XX
Dt      02-AUG-1999 (first entry)
XX
DE      Merozoite surface protein MSP-1-42 modified cDNA.
XX
KM      MSP-1: merozoite surface protein; malaria; vaccine;
KM      protein engineering; protein expression; codon usage;
KM      Transgenic animal; mutant; ss.
OS      Plasmodium falciparum.
OS      Synthetic.
XX
FH      Key      Location/Qualifiers
FT      CDS      1..1131
FT      sig_peptide 1..45
FT      mat_peptide 46..1128
FT      /*tag- b
FT      /*tag- c

WO9920766-A2.
XX
PD      29-APR-1999.
XX
PF      20-OCT-1998; 98WO-US22225.
XX
PR      15-MAY-1998; 98US-0085649.
PR      20-OCT-1997; 97US-0062592.
XX
PA      (GEN2 ) GENZYME TRANSGENICS CORP.
XX
PI      Chen LH, Meade H;
XX
DR      WPI; 1999-302742/25.
DR      P-PSDB; AAY05834.
XX
PT      New modified recombinant nucleic acid sequences useful for producing
PT      malarial DNA vaccine
XX
PS      Claim 12: Fig 11; 43pp; English.
XX
CC      This novel, modified nucleic acid encodes a 42 kDa C-terminal
CC      part (see AAY05834) of malaria merozoite surface protein MSP-1
CC      (MSP-1-42), an important target for the development of a vaccine
CC      against Plasmodium falciparum. The nucleic acid sequence has been
CC      modified compared to the native sequence of MSP-1-42 (see AAK25587)
CC      such that the AT content has been reduced and 10 mRNA instability
CC      motifs eliminated while maintaining the protein amino acid sequence.
CC      In addition, a sequence encoding a 15-amino acid beta-casein signal
CC      peptide has been added to the 5' end of the sequence, and N262Q
CC      and N181Q mutations have been introduced to eliminate
CC      N-glycosylation sites. These alterations allow MSP-1-42 to be
CC      expressed in the mammary gland (i.e. milk) of transgenic mice. The
CC      invention provides modified recombinant nucleic acid sequences and
CC      methods for increasing the mRNA levels and protein expression of
CC      proteins that are difficult to express in cell culture systems,

```

```

CC      mammalian cell culture systems or in transgenic animals. The
CC      preferred difficult protein candidates for expression are those
CC      derived from lower organisms such as parasites, bacteria and
CC      viruses that have DNA coding sequences of high AT content or which
CC      have mRNA instability motifs or rare codons relative to the
CC      recombinant expression system to be used. The invention allows
CC      expression of MSP-1 in the milk of transgenic animals, and also
CC      provides a DNA vaccine comprising a vector containing the altered
CC      MSP-1-42 sequence.
CC
XX
SQ      Sequence 1140 BP; 353 A; 282 C; 290 G; 215 T; 0 other;
Query Match      15.0%; Score 742; DB 20; Length 1140;
Best Local Similarity 80.8%; Pred. No. 3.3e-170;
Matches 865; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
Oy      3791 CCGGTAGGCTGTCACCTCTCGTATGATTAATGATTCGCAAAATGAGAGAAAT 3850
Db      38 CCAATGCGCGTCACTCCCTCGTATGATTAATGATTCGCAAAATGAGAGAAAT 97
Oy      3851 ACGAAGTGTCTATTCGAAACCTCTGCGACGCGCTTATAGTCTTCAGAAACAGCTGG 3910
Db      98 ACGAGTGTCTATTCGAAACCTCTGCGACGCGCTTATAGTCTTCAGAAACAGCTGG 157
Oy      3911 AGAATAACGTGATGACCTTAAATGTCACGCTGAAGACATTCGAAACGCGCTTTATA 3970
Db      158 AGAACACGTGATGACCTTAAATGTCACGCTGAAGACATTCGAAACGCGCTTTATA 217
Oy      3971 AGAGAGAAATTTCAAGAACGCTTGAGAGACGATTCCTTAAAGACCTGACCT 4030
Db      218 AGAGGAGAAATTTCAAGAACGCTTGAGAGACGATTCCTTAAAGACCTGACCT 277
Oy      4031 CCTCTAATCAGTTGTCAAGAACCTATCAAGTTCTCTCAATTAAGAGAGGATTAAT 4090
Db      278 GCAGCAACTGCTGCTCAAGAACCTATCAAGTTCTCTCAATTAAGAGAGGATTAAT 337
Oy      4091 TTCTGTCTACTTCACTATATCAAGAACCTATCAAGTTCTCTCAATTAAGAGAGG 4150
Db      338 TTCTGTCTACTTCACTATATCAAGAACCTATCAAGTTCTCTCAATTAAGAGAGG 397
Oy      4151 ATGTGCTGGGTATTAACAAGATCTGAGCGAAATTAACAAGTGTGACCTTACTATTA 4210
Db      398 ATGTGCTGGGTATTAACAAGATCTGAGCGAAATTAACAAGTGTGACCTTACTATTA 457
Oy      4211 AAAAGTATATCAACGATTAACAAGCGAGATGAATGAATTAATTCGCTTCTTAATACA 4270
Db      458 AGAAGTATATCAACGATTAACAAGCGAGATGAATGAATTAATTCGCTTCTTAATACA 517
Oy      4271 TCGAAACCTGTACAGACGATGAAGCAAAATTCGACCTTCTGATTCACCTGGAG 4330
Db      518 TCGAAACCTGTACAGACGATGAAGCAAAATTCGACCTTCTGATTCACCTGGAG 577
Oy      4331 CCAAGTCTCTCACTATTAACAAGACGATGAAGCAAAATTCGACCTTCTGATTCACCT 4390
Db      578 CCAAGTCTCTCACTATTAACAAGACGATGAAGCAAAATTCGACCTTCTGATTCACCT 637
Oy      4391 ACTACCTCAAAACATCAAGACGATGAAGCAAAATTCGACCTTCTGATTCACCT 4450
Db      638 ATTACCTCAAAACATCAAGACGATGAAGCAAAATTCGACCTTCTGATTCACCT 697
Oy      4451 GAATTCGAGACCTGTCTACCGATTATTAACAACAATTCGACCTTCTGATTCACCT 4510
Db      698 GAATTCGAGACCTGTCTACCGATTATTAACAACAATTCGACCTTCTGATTCACCT 757
Oy      4511 CTGCACTGGTTCGAAACCTGCGCAAAACATTCGATTCGATTCGATTCGATTCGAT 4570
Db      758 CCGGAATGCTTTCGAAACCTGCGCAAAACATTCGATTCGATTCGATTCGATTCGAT 817
Oy      4571 TGCAGGAGATGCTGAACATTCGCAACACCAATGCTGAAGAAACAGTCCCAAGATA 4630
Db      818 TGCAGGAGATGCTGAACATTCGCAACACCAATGCTGAAGAAACAGTCCCAAGATA 877
Oy      4631 GCGGCTGTTTCAGGATTCGCAACACCAATGCTGAAGAAACAGTCCCAAGATA 4690

```


XX	AAK56009;
DT	31-AUG-1999 (first entry)
XX	
DE	Merozoite surface protein MSP-1-42 cDNA.
KM	MSP-1; merozoite surface protein; malaria; vaccine;
KW	protein engineering; protein expression; codon usage;
XX	transgenic animal; ss.
OS	Plasmodium falciparum.
XX	
FT	Key Location/Qualifiers
CDS	/..1086
	+tag= a
XX	
PN	WO9920774-A2.
PD	
XX	29-Apr-1999.
PF	20-OCT-1998; 98WO-US22226.
XX	
PR	15-MAY-1998; 98US-0085649.
PR	20-OCT-1997; 97US-0062592.
XX	
PA	(GENZ) GENZYME TRANSGENICS CORP.
XX	
PI	Chen LH, Meade H;
DR	WPI: 1999-288313/24.
DR	P-PDSB; AAY09373.
XX	
PT	Modified malarial protein for use in anti-malarial vaccines
PS	Example; Fig 2; 35pp; English.
CC	This nucleic acid encodes a 42 kDa C-terminal portion (see AAY09373)
CC	of malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC	important target for the development of a vaccine against
CC	Plasmodium falciparum. The 3' end of the sequence has been
CC	modified to include a 6xHis tag. The nucleic acid sequence has
CC	been modified (see AAK56008) according to a method of the invention
CC	in order to improve expression in mammalian host cells and in
CC	transgenic animals. In the modified coding sequence, 306 nucleotide
CC	positions have been replaced to lower the AT content (from 76 to
CC	49.7%) and to eliminate 10 mRNA instability motifs. The encoded
CC	amino acid sequence is unaltered. In another modified sequence
CC	(see AAK56021), a signal peptide sequence has been added and two
CC	N-glycosylation sites eliminated. The invention allows expression
CC	of MSP-1-42 protein in the milk of transgenic animals, and also
CC	provides a DNA vaccine comprising a vector containing the altered
CC	MSP-1-42 sequence.
XX	
Sequence	1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;

Query Match	12.7%;	Score 625.2;	DB 20;	Length 1088;
Best Local Similarity	74.3%;	Pred. No. 8.7e-142;		
Matches 789;	Conservative	0;	Mismatches 273;	Indels 0;
				Gaps 0;

[illegible]

D	b	181	AATTTCAAAATGTTTAGATCAGATTAACTCCATATAAAGATTAAACATCAAGAAT	240
O	y	4039	TACGTGTGCAGAGACCATTACAGTCCATGCCATAAAGAGAGGAATTAATTCGTCT	4098
D	b	241	TATGTGTGCAGAGATCCATATTAATTTCTTAATAAAGAAAAAGATTAATTCCTTAAGC	300
O	y	4099	AGTTACAACATATATCAAGAGCTCCATCGACACCGCATATCAATTCGGCTAAATGATGTGCT	4158
D	b	301	AGTTATATTAATTAATTATAGGATTCATATAGATACGGAATATAAATTTTGCAAAATGATGTCTT	360
O	y	4159	GCGTATTTCAGAGATCCCGACCGAAAAATACAGTCTGACTGACTCTATTAAAAAGAT	4218
D	b	361	GGATATTATTAATAAATATTTATCCGAAAAATTAATTCAGATTGATTCATTAATAAATAT	420
O	y	4219	ATCACACATTAAGCAGGCGCGAATGAGAAAAATATCTCCCTCGTAAATPAACATCGAACC	4278
D	b	421	ATCACACGACAAACAGAGTGAATGAAATGAGAAATACCTCCCTTTTAAACAATTTGAGACC	480
O	y	4279	CTGTACACAGACAGTGAACGACAAAATGACCTCTTGATATTCACCTGGAGCCCAAGTC	4338
D	b	481	TTATATATAAACAGATTATGATTAATAAATTTGATTTTGTATTAATTCATTTAGAAACAAATCTT	540
O	y	4339	CTCAACTATCTACTACGAGAGACCAATGTGGAAAGTTAAATFCAAGAGCTGAATCACTC	4398
D	b	541	CTTAATTTATACATTTGTGAGAAATCAAACGTRTGAAGTTAAATTAAGAACTTATATCTACTA	600
O	y	4399	AAAACAATCCAAGACANGCTGGCAGATTTCCAGAAAAATTAACAATTTGCTGGAAATTGCA	4458
D	b	601	AAAACAATTCAGACCAATATTGGCAGATTTTTAAAAAAAATTAACAATTTGTTGGAAATTCCT	660
O	y	4459	GACCTGTCTCCCGATTATTTAAACCAACAATCTCTGACCAAGTTTCTGCCACTGGCATG	4518
D	b	661	GATTTATTCACAGATTTTAAACCAATTAATTAATTTTACAAAAGTCTTATAGTACAGATATG	720
O	y	4519	GTGTGTCGAAAACTCGCCAAACAGCTGTCGACATCTGCTCGACGCGCACTCGAGGC	4578
D	b	721	GTTTTGAATAATCTTGCTATAAACCGTTTATCTAATTTACTTGATGAGAAACTTGCAGAGT	780
O	y	4579	ATGCTGAACATCTCCCAGCGACCAATGCGTAGAAGAAACAGTGGCCCCAGATGAGGGCTGT	4638
D	b	781	ATGTTAATACATTTTACACACACCATGTGTRAAAAAACATATGCCACAAAATTTCTGATGT	840
O	y	4639	TTTCAGGACTTCGACGAGCCCGGAAGATGCAACTGTCTCCTGAACTACAAACAGAAAGGA	4698
D	b	841	TTTCAGACATTTAGATGTAAGAAAGAAAGATGTAATGTTTATTAATTAACAAACAGAAAGST	900
O	y	4699	GATATAGTGCCTGGAGAAACCCAAACCTACTGCAATGAAAGAAACAAATGGCGGTGTACACC	4758
D	b	901	GATTAATGTGTGTGAATAATCCCAAACTCTACTGTAAAGAAAAATATGTGTGATGTATCA	960
O	y	4759	GATGCTAAATGACACCGAGAGAGACAGGGCTTAAACGAAAGAAATCATCATGCGAGTGT	4818
D	b	961	GATGCCAAATGTACCGAAGAAAGATTCAGGTAGACGGAACGAAAGAAATTCATATGTGAATGT	1020
O	y	4819	ACTAAGCCGCACTCTATCCACTCTTCGACGGGAGTTTTTGC	4860
D	b	1021	ACTAACCCTGATTTCTATACCATTTTCGATGATTTTCTGC	1062
<hr/>				
RESULT 14				
AAK25587				
ID	AAK25587 standard; cDNA; 1088 BP.			
XX	AAK25587;			
XX	02-AUG-1999 (first entry)			
DT	Merozoite surface protein MSP-1-42 cDNA.			
XX	MSP-1; merozoite surface protein; malaria; vaccine;			
KW	protein engineering; protein expression; codon usage;			
KW	transgenic animal; ss.			
XX				

RESULT 14
AAAX25587
ID AAAX25587 standard; cDNA; 1088 BP.
XX
AC AAX25587;
XX
DT 02-AUG-1999 (first entry)
XX
DE Merozoite surface protein MSP-I-42 cDNA.
XX
KM MSP-1; merozoite surface protein; malaria; vaccine; protein engineering; protein expression; codon usage transgenic animal; ss.
XX

OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..1086
/*tag= a
XX
PN MO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI: 1999-302742/25.
DR P-PSDB: AAY05833.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Disclosure: Fig 2; 43pp; English.
XX
CC This nucleic acid encodes a 42 kDa C-terminal portion (see AAY05833)
CC of malarial merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The 3' end of the sequence has been
CC modified to include a 6xHis tag. The nucleic acid sequence has
CC been modified (see AAX25586) according to a method of the invention
CC in order to improve expression in mammalian host cells and in
CC transgenic animals. In the modified coding sequence, 306 nucleotide
CC positions have been replaced to lower the AT content (from 76 to
CC 49.7%) and to eliminate 10 mRNA instability motifs. The encoded
CC amino acid sequence is unaltered. In another modified sequence
CC (see AAX25593), a signal peptide sequence has been added and two
CC N-glycosylation sites eliminated. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, mammalian cell culture systems
CC or in transgenic animals. The preferred difficult protein
CC candidates for expression are those derived from lower organisms
CC such as parasites, bacteria and viruses that have DNA coding
CC sequences of high AT content or which have mRNA instability motifs
CC or rare codons relative to the recombinant expression system to be
CC used. The invention allows expression of MSP-1 protein in the milk
CC of transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX
SQ Sequence 1088 BP; 454 A; 139 C; 150 G; 345 T; 0 other;
Query Match 12.7%; Score 625.2; DB 20; Length 1088;
Best Local Similarity 74.3%; Pred. No. 8,7e-142;
Matches 789; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 3799 GCTGTCACTCCTCCGTCGATGATACATTTCTGCCAAATCGAGACGATACGAAGTG 3858
DB 1 GCAGTAACCTCTCCGTAATGATACACTTCTTCAAAATGGAATACAGAGTT 60
QY 3859 CTCTATCTGAACCTCTGCGAGCGCTATAGTCTCAAGAAACAGCTGAGAAATAC 3918
DB 61 TTATATTTAAACCTTTGACGAGCTTTATAGAACTTTAAAAAACATTTAGAAAATAC 120
QY 3919 GTGATGACCTTCAATGTCAGTGAAGACATTTGAAACACCGCTTTAATAAGAGAA 3978
DB 121 GTTATGACATTTAATGTAATGTAAGATATTTTAAATTCACGATTAATTAACGTGAA 180
QY 3979 AATTTCAGAACGCTTGGAGAGCACTTGATCCCTATAAGACCTACCTCCTAAC 4038
DB 181 AATTTCAAAATGTTTAGAATCAGATTTAATTCATTAATAAGATTAAACATCAAGTAAT 240

QY 4039 TACGTTGCAAGGACCATATACAGTTCCATATAAGAGAGGATAAATTCGTCT 4098
DB 241 TATGTTGCCAAGATCCATATTAATTTCTTAATTAAGAAAAAGATATAATTTCTTAAGC 300
QY 4099 AGTTACAACTATATACAGAGACTCCATCGACACCGATATCAATTTGCTAATGATGCTG 4158
DB 301 AGTTAATATTTAATTAAGAGATTCATAGATACGATATTAATTTTGCAAATGCTTCTT 360
QY 4159 GGGTATTTACAGATCTCGAGGAAAAATACAGTTGACTGACTTATTTAAAAAGTAT 4218
DB 361 GGATATTTTAAAAATATTTATCCGAAAAATATTAATGATTAAGATTCATTTAAAAAATAT 420
QY 4219 ATCAACGATTAAGCAGGGGAGAGAAAAATATCTGCCCTCCCTGTAATACATGCAACC 4278
DB 421 ATCAACGACAAACAAAGTGAATAGAAATGAAATCTCCCTTTTAAACATATTTAGACC 480
QY 4279 CTGTACAAGACAGTGAACGACAAATTCGACCTCTTCTGTAATTCACCTGGAGCCAAAGTTC 4338
DB 481 TTATATTAAGAGTAAATGATTAATTTGATTTGTAATTCATTTGAAGCAAAAGTT 540
QY 4339 CTCAACTTACTTTACGAGAGACGCAATGTGAAATTAATCAAGAGACTGAACTACTCTC 4398
DB 541 CTAAATTTACATATATGAGAAATCAAAACGTAGAGTTAAATAAAAAAGAACTTATTTACTT 600
QY 4399 AAAACAAATCCAAAGACAGCTGGCAGATTTCAGAAAAATTAACAATTTGCTGGAAATTCGA 4458
DB 601 AAAACAAATTCAGACAAATTTGGCAGATTTTAATAAAAAATTAACAATTTGCTGGAAATTCGT 660
QY 4459 GACCTGTCTACCGATTATTAACCAACAATCTCTGACCAAGTTTCTGCACTGGCAATG 4518
DB 661 GATTTATCAACAGATTATTAACCATTAATTAATTTGACAAAGTTCTTAGTACAGATG 720
QY 4519 GTGTTGCAAAACCTGCGCAAAACAGTGTGACAAATCTGCTGACAGCACTGCAAGGCG 4578
DB 721 GTTTTGAATAATCTTGCTAAACCCGTTTATCTAATTTATTTGATGAGAAACCTTGCAAGGT 780
QY 4579 ATGCGAATCTCCGACCAATCGTGAAGAAACAGTGCCGCCCAATTAAGCGGCTGT 4638
DB 781 ATGTTAAACATTTTCACACCAACCAATGCGTTAAAAAACATATGTCACAAATTTCTGATGT 840
QY 4639 TTCAGCATCTGAGAGAGCGGAGAGAGTGCAGAGTGTCTCTGAACTACAAACAGAGGA 4698
DB 841 TTCAACATCTTAGAAGAAAGAGAAATGTAAATGTTTATTAATTAACAACAAAGAGT 900
QY 4699 GATAGTCTGCGGAGAACCCCAACCTCTACCTGCAATGAAACAAATGGGGGTGTAGCC 4758
DB 901 GATTAATGTGTAAGAAATCCAAATCTTACTGTAAAGAAATTAATGATGATGATGATCA 960
QY 4759 GATGCTAATGCAACCGAGAGACAGCGCTCTTAACGGAAGAAATCAATGCAATGCTGT 4818
DB 961 GATGCCAATGTACCGAAGAAAGATTCAGTAGCAACGGAAGAAATCAATGATGATGT 1020
QY 4819 ACTAAGCCCGACTCTATTCACCTTTCGACGGAGATTTTTCG 4860
DB 1021 ACTAATCCTGATTTCTTATCCACTTTTCGATGTATTTTCG 1062
RESULT 15
AAC68978 standard; DNA; 786 BP.
XX AAC68978;
XX
XX 27-FEB-2001 (first entry)
XX
XX Merozoite surface protein-133 coding sequence.
XX DE Merozoite surface protein; protazoacide; vaccine; malaria; ss.
XX KW Merozoite surface protein; protazoacide; vaccine; malaria; ss.
XX OS Plasmodium falciparum.
XX PN MO200063245-A2.

CC malaria vaccine, which is used to treat or prevent malaria, caused by
 CC any of the four species of Plasmodium that infect humans.
 XX
 SQ Sequence 1128 BP; 487 A; 149 C; 163 G; 329 T; 0 other:

Query Match 6.5%; Score 322.8; DB 22; Length 1128;
 Best Local Similarity 58.3%; Pred. No. 3,66-68;
 Matches 644; Conservative 0; Mismatches 397; Indels 63; Gaps 2;

```

OY 3817 ATTGTAACATTCTGTCACAAAATGAGAACGATACGAGTGTCTATCTGAAACCTCTG 3876
    || || || || || || || || || || || || || || || || || || || || || ||
DB 19 ATGTAATATCTCTCTCGAGATTGAAATGATGATATATATTTAAACCTTTA 78

OY 3877 GCAGCGCTATAGTCTCTCAAGAAACAGCGAGATTAACGTCATGACCTTCAATGTC 3936
    || || || || || || || || || || || || || || || || || || || || || ||
DB 79 GCTGAGATATATAGAGCTTTAAAAACAAATTGAAAAACATTTTACATTTTAAATTA 138

OY 3937 AACGTGAAGACATTCAGACAGCGCTTTAATAAGAGAAATTTGACAGACCTCTTG 3996
    || || || || || || || || || || || || || || || || || || || || || ||
DB 139 AATTGAAGATATCTTAATTCACGTCTTAAGAAACGAAATATTTCTTACATGATTA 198

OY 3997 GAGAGCGACTGATTCCTATTAAGACCTGACCTCTCTACTACCTGTCAGAGACCA 4056
    || || || || || || || || || || || || || || || || || || || || || ||
DB 199 GAATCTGATTAATGCAATTTAAACATATATCCCAATGATATCATTTATGAAATGCA 258

OY 4057 TACAAGTCTCTCAATTAAGAGAGAGGATTAATTTCTGTCTAGTTACACTATATCAAG 4116
    || || || || || || || || || || || || || || || || || || || || || ||
DB 259 TTTAAATATATGATTCGAAACAAAAACACACTTTTAAAAAGTTTACAAATATATTA 318

OY 4117 GACTCCATCGACACCGATATCAATTTGCTATATGATGCTGGGATTTACAGATCTG 4176
    || || || || || || || || || || || || || || || || || || || || || ||
DB 319 GAATCAGTAGAAAAATGATTAATTAATTTGACAGAGGATATAGTTATATGAAAAAGTT 378

OY 4177 AGCGAAAAATFACAGTCTGACCTTACTCTATTTAAAAATATATACGATTAAG----- 4230
    || || || || || || || || || || || || || || || || || || || || || ||
DB 379 TTAGGAAATATTAAGATGATTTAGAAATCAATTTAAAAAGTTATCAAGAGAAAAAGAG 438

OY 4231 -----CAAGC 4236
    || || || || || || || || || || || || || || || || || || || || || ||
DB 439 AAGTTCATCATCACCAACAACACCTCGTCACCAAGAAAAACAGACGACAAAAAG 498

OY 4237 GAGAAATGAAAAATATCTGCTCTGTAATAACATGAAACCTGTACAGACAGTGAAC 4296
    || || || || || || || || || || || || || || || || || || || || || ||
DB 499 AAGGAAAGTAAAGTCTCCATTTTAAACAAACATTTGACACTTATCAATTAAGTT 558

OY 4297 GACAAATATGACCTCTGTATATGACGAGGCAAGGTCCTCACTATCTTACGAG 4356
    || || || || || || || || || || || || || || || || || || || || || ||
DB 559 AATTAATATGACGATTAATTAAGGCAAGATTAACGATTTGATGATTAAGTGA 618

OY 4357 AAGAGCAATGTGAGAAATTAATCAAGAGCTGAATCACTCAAAACAATCCAGACAAG 4416
    || || || || || || || || || || || || || || || || || || || || || ||
DB 619 AAGATGAGAGCAATGTTAAATTAATTAAGTGAATTAAGCAATTCATGATACAAA 678

OY 4417 CTGGCAGATTTCAAGAAAAATTAACATTTGCTGGAATTCGACACTGTCTACGATTA 4476
    || || || || || || || || || || || || || || || || || || || || || ||
DB 679 ATAGATCTTTTAAAAAACCTTACACTTCGAAAGCAATTTAAAAATGATAATGATAT 738

OY 4477 AACCAACAATTCCTGACCAAGTTTCTGTCACACTGTCATGCTGTGGAANAACCTGCC 4536
    || || || || || || || || || || || || || || || || || || || || || ||
DB 739 ACGAAAAAAGATATCTTGGCAAAATTAATTAAGTGAAGATTAAGTGAAGATTAAGT 795

OY 4537 AAAACAGTGTGAGCAATCTGCTGACGCGCAACCTGACGGGCTCTACATCTCCAG 4596
    || || || || || || || || || || || || || || || || || || || || || ||
DB 796 AATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 855

OY 4597 CACCAATCGTGAGAGAAACAGTCCCGCAGAAATGACGCTGTTTCAGCATCTGACGAG 4656
    || || || || || || || || || || || || || || || || || || || || || ||
DB 856 CACCAATCGTGAGAGAAACAGTCCCGCAGAAATGACGCTGTTTCAGCATCTGACGAG 915

OY 4657 CGCGAAGGTGCAAGTGTCTCTGAACTAACAAACAAGAGAGATTAAGTGTGAGAGAC 4716
    || || || || || || || || || || || || || || || || || || || || || ||
DB 916 AGAGAGAGATTAATGATTTAATTAATTAACAACAAGAGGATGATTAATGATGATTAAT 975
  
```

```

OY 4717 CCAACCTACCTGCATGAATAAACAATGAGGCGTGTGAGCGCCGATGTAATGACACGAG 4776
    || || || || || || || || || || || || || || || || || || || || || ||
DB 976 CCAAAATCTACTTGTAGAGAAATATATGATGATGATGATGATGATGATGATGATGAT 1035

OY 4777 GAGACAGCGGCTCTAACGAGAAATATCAATGTCGATGATGATGATGATGATGATGAT 4836
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1036 GAGATTCAGGTAGAGACAGAGAAATATCAATGATGATGATGATGATGATGATGATGAT 1095

OY 4837 CCACCTTCGACGCGGATTTTTCG 4860
    || || || || || || || || || || || || || || || || || || || || || ||
DB 1096 CCACCTTCGACGCGGATTTTTCG 1119
  
```

RESULT 17

ID AN81148 standard; DNA; 660 BP.

AN81148;

11-NOV-1990 (first entry)

DNA sequence encoding polypeptide p190-1.

Polypeptide p190-1; P.falciparum; merozoite; vaccine.

Plasmodium falciparum.

Key Location/Qualifiers

FT CDS 1..660

EP283829-A.

28-SEP-1988.

08-MAR-1988; 88EP-0103564.

19-MAR-1987; 87GB-0006599.

(HOFF) F. HOFFMANN-LA ROCHE & CO.

Ulrich C, Gentz H, Takacs B;

WPI; 1988-272339/39.

P-PSDB; AAP80544.

New immunogenic polypeptides - derived from plasmodium falciparum 190 kD surface antigen precursor, useful in vaccines, and encoding DNA sequences

Disclosure; pp; German.

The DNA encodes a region of a polypeptide. The region contains at least one epitope of the 190KD precursor of the major merozoite surface antigen of P.falciparum.

Sequence 660 BP; 281 A; 84 C; 102 G; 193 T; 0 other;

Query Match 6.3%; Score 308.8; DB 9; Length 660;

Best Local Similarity 74.1%; Pred. No. 7,1e-65;

Matches 391; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```

OY 475 ACACGTGTGATTAATGATGCTTCAAAATCTGATTAAGGCTTAAGAGAGATCAAT 534
    || || || || || || || || || || || || || || || || || || || || || ||
DB 25 ACTTGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 84

OY 535 GAATCTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 594
    || || || || || || || || || || || || || || || || || || || || || ||
DB 85 GAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 144

OY 595 TGCGCAATGACTATTTGCAATTTCCATTTGAAGATGACAGCCAGCATGTTGAC 654
    || || || || || || || || || || || || || || || || || || || || || ||
DB 145 TGCTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 204
  
```


QY	655	GTATTGAAGAGTTGGCTTGGATTTATCGAAGCCCTCGCAACATCAAGGAAACATGTG	714
Db	205	GTACTTAAAAAACTTGtGTGGATATGAAAAACATTAGCAATATTAAATAAATGTA	264
QY	715	GGAAAGATGGAAGATTATATATAAAAAGATATAGAAAGACCATCGAAGATTAAAGAGCTG	774
Db	265	GGAAAAATGGAAGATTTACATTAAAAAAATATAAAAAACCATAGAAATATTAATGAAATT	324
QY	775	ATCGAAGATCCAAAAAGACCATTAGACAAAAATTAAGATGCAACCAAGAGAGAAAG	834
Db	325	ATTGAGAGAACTTAAGAAAAACATTTGATTAATAATTAAGATGAACCTTAAGAAAGAA	384
QY	835	AAGAAATTGACCAAGGCCAGTACGACCTGTCCATCTATAACAACAGCTTGAAGAGCC	894
Db	385	AAAAATATATCCAAAGCTCATATATGATCTCTTTTATTACATAAACAATTAGAAAGCA	444
QY	895	CATAACCTATCAGCGTACTGGAAGAGCGCATATACACCCCTCAAGAAGATGAAATATC	954
Db	445	CATATATTAAATACCGTTTGAAGAAAAAGTATTACACTTTAAAAAAAATATGAAACATT	504
QY	955	AAAGAACTGTCGACAAGATTATGAAGATTAAGAACTCCGCGCAGCC	1002
Db	505	AAGGATTACTTGATTAAGATTAATGAAGAAATTAATATCCCCACCGGCGC	552

```
RESULT 18
AAN81151
ID AAN81151 standard; DNA; 1068 BP.
```

AC AAN81151;

DT 11-NOV-1990 (first entry)

DNA sequence encoding polypeptide p190-3.

Polypeptide p190-3; *P. falciparum*; merozoite; vaccine

OS *Plasmodium falciparum*.

FH	key	Location/qualifiers
ET	CDC	1 1068

/*tag= a

PN EP283829-A.
xy

PD 28-SEP-1988
YY

PF 08-MAR-1988; 88EP-0103564.
XX

PR 19-MAR-1987; 8/GB-0006599-
XX XX

PA (HOFF) F. HOFFMANN-LA ROCHE & CO.

Ulrich C, Gentz H, Takacs B,
PI
XX

DR WPL; 1988-2/2339/39.
DB B-PEND; ADP80547

New immunogenic r

surface antigen precursor, useful in vaccines, and encoding DNA sequences

PS Disclosure; ; cpp; German.
XX

CC The DNA encodes a region of a polypeptide. The region contains at least
CC one epitope of the 190kd precursor of the major merocystite surface antigen

CC 01 P. talc. parum.
YY

sequence 1068 BP; 4/1 A; 129 C; 161 G; 307 H; 0 other;

Query Match	6.3%;	Score 308.8;	DB 9;	Length 1068;
Best Local Similarity	74.1%;	Pred. No. 9e-65;		
Matches 391; Conservative	0;	Mismatches 137;	Indels 0;	Gaps 0

QY	475	ACACGTGTGATTAACATTCATTCAGCGCTTCAAAATATCGATTGATTCGAGAAATCAAT	534
Db	433	ACTTTGTGTATTAATTCATTCAGCTTCAAAATTTAATTTGATGATTTGAAGAAATTAAT	492
QY	535	GAACCTCTGATCAAGTTGAATTTCTTACTTGATCGACTTGAGGGCCAAACTGATGACGTT	594
Db	493	GAATTAATTAATTAATTAACCTTTAATTTTGATTTAATTAAGAGCAAAATTTAAATTAAGTA	552
QY	595	TGGCCCATGACATTTGTGCAAAATTCATTCGAATTTGAAGATCAGAGCAACGATTTGGAC	654
Db	553	TGTCTATATGATTTATGTGCAPATACCTTTCAATCTTTAAATTTGCTGCAATTTGATGAC	612
QY	655	GTATTGAAGAAGTTGGTCTTCGATATCGAAGCCCTCTCGACAACATCAAGACAAATGTG	714
Db	613	GTACTTAAAAAACTGTGTGTCGATATAGAAAACCATTTAGACATATTTAAAGATAAATGTA	672
QY	715	GGAAAGATGGAAAGTTATATTTAAAAAGAAATTAAGAAACCATCGAAGACATTAAGACTG	774
Db	673	GGAAAAATGGAAGATTACATTTAAAAAAAATAAAAAACCATAGAAATATTAATGAAGTTA	732
QY	775	ATCGAGACATCCAAAAGACCATAGACAAAATAAGATAGCAACCAAGAGAGAAAG	834
Db	733	ATTGAGAAAGATAGAAAACATTTGATTAATAAATTAAGATGCACTTAAGAAGAGAAAA	792
QY	835	AAGAAGTTGTACGAGGCCAGTACGACCTGTCTATTAACAACAACGTTGAAGAAAGCC	894
Db	793	AAAAAATTAATACCAAGCTCAATATGATCTTTTATTTTACATAAACAATTTAGAGAAAGCA	852
QY	895	CATTAACCTCATCACCGTATCTGGAAGACGCGATACACCCCAAGAAGATGAATAATTC	954
Db	853	CATATTTAATTAAGCGTTTGTGAAAAACGATTTGACACTTTAAAAAAAATGAATAACATT	912
QY	955	AAAGAACTGCTCGACAGATTAATGAATTTAAGATCTCTCGCCAGCC	1002
Db	913	AAGGAATTAATCTGATTAAGATTAAGAAATTTAAATATCCCATCCGCGCC	960

RESULT 19	
AAT80403	
ID AAT80403	standard; DNA; 354 BP.

AC AAT80403;

DT 25-MAR-1998 (first entry)

DE PFMSP1 (p19)A coding sequence.

KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;

XX

OS Synthetic.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

$$\frac{FT}{a} / *tag=$$

FT /note="sequence contains 2 in frame stop codons at

```
FT misc_feature 1..285 |
```

FT /note= "derived from P. falciparum MSP1 p19 sequence

FT

/*tag= c

XX

XX

XX

XX

100


```

OY 337 AGAACCAATCATCTGACAACTCTAGCAATTCGAGCCAACTCCTACCGACCTCAAG 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 AGTCAAAATCCTTCAGATATATCAAGTATGATGCAATCTTACCTATTTAAAA 192
OY 337 CACCGAGTGAAGAACTATCTCTCACTATCAAGAGCTGAAGTACCACAGTTGTTGAC 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 CATAGAGTTCAAAATTAATCTGTTCACTATTTAAAGAACTCAATATCCGAACTCTTTGAT 252
OY 457 CTCATATATCATATGCTACACTGTGTATATACATTCATGCTTCAAAATCTGATGAC 516
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 TTAACCAATCATATGTATTAATCTTGTGATATATATCAATGTTCAAAATTTAATTTGAT 312
OY 517 GGTTCAGAGAGATCAATGAATCCTGTAACAGTTGAATTTCTACTGACTTGTAAAG 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 GGATATGAAGAAATTAATGAATTAATATATATTAATTAATCTTATATGATTTAATAGA 372
OY 577 GCCAACTGAATGACGTTTGGCCCAATGACTATTTGTCAAATTCATTCATTTGAAGATC 636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 GCCAAATTAATGATGATGATGCTCAATAGTATGTCAAAATACCTTCAATCTTAAAT 432
OY 637 AGAGCAAGAGTGTGAGTATGAAAGTGTGCTTCCGATTCGCAAGCTCTCGAC 696
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 CTGCAAAATGAATTAAGTACTTAATAAAATTTGTTGATATGAAGAAACCATTAGAC 492
OY 697 AACATCAAGACATGTGGAGAGATGAGATTAATTAATAAGATTAAGAAGACCATC 756
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 493 AATATTAAGATTAATGAGAAAAATGAGATTTACATTAATAAAATTAACAACCATTA 552
OY 757 GAGAACATTAAGAGCTGATCGAAGAAATCCAAAAAGACCATTAAGAAAAATTAAGATGA 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 553 GCATATTAATAATGAATTAATGAAGAGTAAGAAACATTAATCCGGCAACGTGT 612
OY 817 ACCAAGAGAGAG 829
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 GCCATTTGCTGAG 625

RESULT 23
AAN81150
ID AAN81150 standard; DNA; 456 BP.
XX
AC AAN81150;
XX
DT 11-NOV-1990 (first entry)
XX
DE DNA sequence encoding polypeptide p190-2b.
XX
KW Polypeptide p190-2b; P.falciparum; merozoite; vaccine.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..456
FT /tag= a
XX
PN EP283829-A.
XX
PD 28-SEP-1988.
XX
PF 08-MAR-1988; 88EP-0103564.
XX
PR 19-MAR-1987; 87GB-0006599.
XX
PA (HOFF) F. HOFFMANN-LA ROCHE & CO.
XX
PI Ulrich C, Gentz H, Takacs B;
XX
DR WPI: 1988-272339/39.
XX
DR P-PSDB: AAP80546.
XX
PT New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
   surface antigen precursor, useful in vaccines, and encoding DNA sequences

```

```

XX
PS Disclosure; 5 spp; German.
XX
CC The DNA encodes a region of a polypeptide. The region contains at least
CC one epitope of the 190kD precursor of the major merozoite surface antigen
CC of P.falciparum.
XX
SQ Sequence 456 BP; 200 A; 63 C; 70 G; 123 T; 0 other;

Query Match 4.1%; Score 201.8; DB 9; Length 456;
Best Local Similarity 67.5%; Fred. No. 6.5e-39;
Matches 284; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

OY 3214 GCCGAGATCGCCGAGACAGAGAACTCTGGAGAACACCAATTTCTCAACACTAC 3273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16 GGTGAATATGCAAGAACTGAACACATTAAGAAACACAAAATTAATGAAACATTAT 75
OY 3274 AAAGGCTTCGTCAAGTATTAATGCGAGTCTCTCTGAGACTCTCTCGAGAG 3333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 AAAGACTTGTAAATTAATTAATGATGATCATCTCCATTAATAAATTAAAGTGAAGAA 135
OY 3334 AGCATCCAGACCGAGATTAACCGACCTCGAGAACTTAAGTCTCTGTAAGCTC 3393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 TCAATTCAAAACGAAGATTAATTAATGCAAGTTTGAAGAACTTTAAGTAAATTA 195
OY 3394 GAAGCAAGCTGAAGAGACACCTGAACCTGGAAGAGAAAGAACTCAGTACCTCTAGC 3453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 GAAGGAAATTAAGGATTAATTAATTAATTAAGAAAGAAATTAATTAATTAATTA 255
OY 3454 GGACTGCATCACCTGATCGCCGAGCTCAAGAGTCAATTAAGAACAGAACTACCGGC 3513
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 GGTTCATCATTAATTAATGCAAGATTAATAAGAAAGATTAATAAAATTAATTAACAG 315
OY 3514 AATAGCCCAAGCAGAAATTAATACAGATGAATTAACGCACTGGAATTTTACAAGATT 3573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 AATTCCTCAAGCTGAATTAATACGATGTGAACAATGCAATTAATTAATTAATTAAT 375
OY 3574 CTGCTGAAGAGACAGATGTCGACCTGTGATGCTGAATGCTCGACACACTGGAG 3633
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CTCCCAAGAGACAGATGTTGCAACAGTTGTAAGTGAAGATGATCCGTGACCTGCAG 435
OY 3634 C 3634
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 C 436

RESULT 24
AAN81149
ID AAN81149 standard; DNA; 462 BP.
XX
AC AAN81149;
XX
DT 11-NOV-1990 (first entry)
XX
DE DNA sequence encoding polypeptide p190-2a.
XX
KW Polypeptide p190-2a; P.falciparum; merozoite; vaccine.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT CDS 1..462
FT /tag= a
XX
PN EP283829-A.
XX
PD 28-SEP-1988.
XX
PF 08-MAR-1988; 88EP-0103564.
XX
PR 19-MAR-1987; 87GB-0006599.
XX
PA (HOFF) F. HOFFMANN-LA ROCHE & CO.

```



```
XX  Ulrich C, Gentz H, Takacs B;
PI  WPI: 1988-272339/39.
XX  P-PSDB; AAB80545.
DR
XX  New immunogenic polypeptides - derived from Plasmodium falciparum 190 kD
PT  surface antigen precursor, useful in vaccines, and encoding DNA sequences
XX
XX  Disclosure: ; pp; German.
XX
XX  The DNA encodes a region of a polypeptide. The region contains at least
CC  one epitope of the 190kD precursor of the major merozoite surface antigen
CC  of P.falciparum.
XX
SQ  Sequence 462 BP; 200 A; 65 C; 72 G; 125 T; 0 other;

Query Match      4.1%; Score 201.8; DB 9; Length 462;
Best Local Similarity 67.5%; Pred. No. 6.6e-39;
Matches 284; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY  3214 GCGGAGATCGCGAGACAGAGAACACTCTGAGACACACCAAGATTCTTCTCAACACTAC 3273
DB  22 GCTGAATATACAGAACTGAAAACACATGTAGAAAACACAAAATATATTGAAACATTAT 81
QY  3274 AAAGGCTCTCAAGTATATTAATGCGAGCTCTTCTCTGAAGACTCTCTCGAGAG 3333
DB  82 AAAGGACTTGTAAATATTAATGATGATCATCTCCATTAATAAACTTAAGTGAAGAA 141
QY  3334 AGGATCCAGACCGAGATTAATCTACGAGCTCGAGACCTGAGAACTCAAGGCTGTAAAGCTC 3393
DB  142 TCATTTCAAAACAGAGATTAATTAATGCCAGTTTAGAAAACCTTTAAAGTATTAAGTAAATTA 201
QY  3394 GAAGGCAAGCTGAAGACCAACTGAACTGAGAGAGAGAGAGCTCAGCTACTCTTACG 3453
DB  202 GAAGGAAATTAAGGATTAATTAATTAAGGAAAGAAATTAATCATCTTATCAAGA 261
QY  3454 GGAAGCTGATCACTGATGCCGAGCTGAGAGAGTCAATTAAGAACAGAACTACCGGC 3513
DB  262 GGTTCACATCTTAATTAATGCTGATTAAGAAAGTAAATAAAATTAATTAATACAGT 321
QY  3514 AATAGGCCAAGCGAGATTAATTAAGAGCTGAAATTAAGCAGTCAATCTTCAAGAGTTC 3573
DB  322 AATTTCTCAAGCGTAAATTAATTAAGGATTTAATCAATGATTAATCAATTAATTAATTT 381
QY  3574 CTGCTGTAAGAGACAGATGTGCGCACTGTGCTGTCAATCTGCTCGACACACTGAG 3633
DB  382 CTCCAGAGAGAGACAGATGTGCAACAGTTGTAAGTGAAGTGAATGCTCGTGCAGCTGCAG 441
QY  3634 C 3634
DB  442 C 442

RESULT 25
AAC68977
ID  AAC68977 standard; DNA: 333 BP.
XX
XX  AAC68977;
AC
XX  27-FEB-2001 (first entry)
DT
XX  Merozoite surface protein-119 coding sequence.
DE
XX  Merozoite surface protein; protazoacide; vaccine; malaria; ss.
KM
XX  Plasmodium falciparum.
OS
XX  WO200063245-A2.
PN
XX  26-OCT-2000.
PD
XX  20-APR-2000; 2000WO-GB01558.
PF
```

```
XX  20-APR-1999; 99GB-0009072.
PR  13-MAY-1999; 99US-031817.
PR  25-MAY-1999; 99CA-2271451.
XX
XX  (MED-) MEDICAL RES COUNCIL.
PA
XX  Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiybull C;
PI  WPI: 2001-015762/02.
XX  P-PSDB; AAB37609.
DR
XX
XX  Novel variants of the C-terminal fragment of Plasmodium merozoite
PT  surface protein-1, useful as vaccines for treating or preventing
PT  malaria -
XX
XX  Example 5; Fig 15; 126pp; English.
PS
XX
XX  The present invention relates to non-natural variants of a C-terminal
CC  fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC  non-natural variants have reduced affinity for at least 1 antibody
CC  capable of blocking a second antibody that inhibits the proteolytic
CC  cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC  one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC  MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
CC  present invention are useful for immunizing a mammal against malaria, and
CC  can be used to treat malaria. The present sequence is the MSP-119
XX  coding sequence.
SQ  Sequence 333 BP; 107 A; 75 C; 73 G; 78 T; 0 other;

Query Match      4.0%; Score 198.4; DB 22; Length 333;
Best Local Similarity 77.2%; Pred. No. 3.7e-38;
Matches 241; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY  4668 ACCGCGAGGCGATGCTAATCTCCAGACCAAGTCGTGAAGAAACAGTCCGCCAGA 4627
DB  20 ACATTGAAGGTAGACACACAACTTGCCCAACACCATGCTTAAGAACATATGTCACAAA 79
QY  4628 ATAGCGGCTGTTTCAGGACATCTGAGACGAGCGGAGAGTGCATGCTCTCTGAATACA 4687
DB  80 ACTCCGATGTTTCAGACATCTGAGACGAGAGAGAAATGAATGCTGTGTAATACA 139
QY  4688 AACAGAGAGAGATAGTGGTGGAGAACCCAAACCTTACTGATGAATGAACATGGCG 4747
DB  140 AGCGAGGAGGTGATTAATGTTGTTGAGAACCAACCTTACTGTAAGAAACAAACGTG 199
QY  4748 GGTGTGACGCGGATGCTAATTAATGACGAGAGAGACGCGCTTAAGCGAAAGAAATCA 4807
DB  200 GATGCGAGCTGACGCTAATGACCGAGAGAGAGACTGTGTTTAACGGAAGAGATTTA 259
QY  4808 CATGCGAGTGTACTAAGCCCGGACTCTATCCACTCTTGAACGAGGATTTTGTCCAGCT 4867
DB  260 CTTCGGAATGTACTAAGCCAGAGACTTACCTTTGTTGATGAATGATCTGTCTTCCCT 319
QY  4868 CTAAATTTCTCTGG 4879
DB  320 CTAACTAAGTGG 331

RESULT 26
AAT80404
ID  AAT80404 standard; DNA: 387 BP.
XX
XX  AAT80404;
AC
XX  25-MAR-1998 (first entry)
DT
XX  PFMSP1(p19)s coding sequence.
DE
XX  Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KM  Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX
```



```

OS Chimeric - Plasmodium vivax.
XX Chimeric - Plasmodium falciparum.
FH Key Location/Qualifiers
FT CDS 1..387
FT /tag= a
FT /product= pFMSPI(p19)S
FT /note= "sequence contains 2 in frame stop codons at
FT the 3' end"
FT sig_peptide 1..57
FT /tag= b
FT mat_peptide 58..381
FT /tag= c
FT misc_feature 1..96
FT /tag= d
FT /note= "sequence derived from P. vivax MSP1"
FT misc_feature 97..102
FT /tag= e
FT /note= "sequence derived from generated restriction
FT 103..381
FT misc_feature enzyme site"
FT /tag= e
FT /note= "sequence derived from P. falciparum p19 coding
FT sequence"
FT W09730159-A2.
XX 21-AUG-1997.
XX 14-FEB-1997; 97WO-FR00291.
XX 14-FEB-1996; 96FR-0001821.
XX (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX WPI: 1997-425034/39.
XX P-PSDB: AAW22592.
XX Recombinant protein containing Plasmodium merozoite surface
XX protein-1 p42 fragment - useful in antimalarial vaccines, also new
XX antibodies for diagnosis and protein purification
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the nucleotide sequence encoding a chimeric protein comprising
XX nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
XX surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
XX 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX The recombinant protein can be used for the production of anti-malarial
XX vaccines, where the p19 fragment provides a high level of protective
XX immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other:
SQ
Query Match 3.9%; Score 194.8; DB 18; Length 387;
Best Local Similarity 81.3%; Pred. No. 3e-37;
Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

```

DB 221 AGTGGCTGGAGAAACCCCAACCGACCTGTACGAGAACAAGCGCGCTGACGACAGAC 280
QY 4763 CTTAATGCACCGAGAGAGACGGCTCTAAGGAGAAATACATCGAGTGTACTA 4822
DB 281 CCAAATGACACCGGAGGAGACTCGGCGACGACGCAAGGCAAAATCAGCTGTGATGTACA 340
QY 4823 AGCCGACCTCTATCCACTCTTCGACGAGATTTTTCG 4860
DB 341 AACCGACTGTACCCGCTGTTGACGCGCATCTTCTGC 378
RESULT 27
AAT94549
ID AAT94549 standard; DNA; 387 BP.
XX AAT94549;
XX 25-MAR-1998 (first entry)
XX pFMSPI(p19)S coding sequence.
XX DE
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope; ss.
XX OS Chimeric - Plasmodium vivax.
XX Chimeric - Plasmodium falciparum.
XX
XX Key Location/Qualifiers
XX CDS 1..387
XX /tag= a
XX /product= pFMSPI(p19)S
XX /note= "sequence contains 2 in frame stop codons at
XX the 3' end"
XX sig_peptide 1..57
XX /tag= b
XX mat_peptide 58..381
XX /tag= c
XX misc_feature 1..96
XX /tag= d
XX /note= "sequence derived from P. vivax MSP1"
XX /tag= e
XX /note= "sequence derived from generated restriction
XX 103..381
XX misc_feature enzyme site"
XX /tag= e
XX /note= "sequence derived from P. falciparum p19 coding
XX sequence"
XX W09730158-A2.
XX 21-AUG-1997.
XX 14-FEB-1997; 97WO-FR00290.
XX 14-FEB-1996; 96FR-0001822.
XX (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX WPI: 1997-425033/39.
XX P-PSDB: AAW36102.
XX Recombinant protein containing the merozoite surface protein-1 p19
XX fragment - useful in anti-malarial vaccines, diagnosis and protein
XX purification
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the nucleotide sequence encoding a chimeric protein comprising

```


CC nucleotides encoding amino acids 1-32 of the Plasmodium vivax merozoite
 CC surface protein 1 (MSP1) linked to the nucleotide sequence encoding the
 CC 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.

XX Sequence 387 BP; 116 A; 94 C; 96 G; 81 T; 0 other;

Query Match 3.9%; Score 194.8; DB 18; Length 387;
 Best Local Similarity 81.3%; Pred. No. 3e-37;
 Matches 226; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 4583 TGAACATCTCCAGACCAATGGCTGAAGAAGATGCCCCAGAAATAGCGCTGTTTCA 4642
 DB 101 TCACATCTCGACAGCAACATGGCTGAAAAACATGCTCCGAGAACTGGCTGTTTCA 160
 OY 4643 GGCATCTGGACGAGCGGGAAGATGCAAGTGTCTCTGAACTACAAACAGAAAGAGATA 4702
 DB 161 GACACTTGGACGAGAGAGAGAGTGAATGTCTGTGAACACAAACAGAGGCGGACAA 220
 OY 4703 AGTGGCTGGAGAACCCAAACCTACCTGCAATGAAAAAATGGGGGTGTAGCGCGATG 4762
 DB 221 AGTGGCTGGAGAACCCCAACCTGCTACGAGAAACAGCGGCTGTAGCGAGACG 280
 OY 4763 CTAATATGACCGAGAGAGAGAGCGGCTTAACGAAAGAAATACATGCGAGTGTACTA 4822
 DB 281 CCAATATGACCGAGAGAGAGAGTGTGAGCAAGCAAAATACAGTGTGAGTGTACCA 340
 OY 4823 AGCCGACCTCTATCCACTCTTGGAGGGGATTTTTCG 4860
 DB 341 AACCGACCTGTACCGCTGTGTGAGGGCATCTTCTGC 378

RESULT 28
 AAD22459
 ID AAD22459 standard; DNA: 309 BP.

AC AAD22459;
 DT 12-FEB-2002 (first entry)

DE Plasmodium falciparum PfMSP1.19 insert in pUC105-01 vector.

KW pUC105-01 vector; immunostimulant; vaccine; immunisation; therapeutic;
 KW C3d; immune response; ds.

OS Plasmodium falciparum.

PN WO200177324-A1.

PD 18-OCT-2001.

PF 09-APR-2001; 2001WO-GB01599.

PR 08-APR-2000; 2000GB-0008582.

PA (ADPR-) ADPROTECH LTD.

PI Steward M, Cox VF;

XX WPI; 2002-010909/01.

PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
 PT occurring protein and comprises a sequence non-identical to naturally
 PT occurring DNA sequence encoding the protein -
 XX Example 13; Page 65; 87p; English.

CC The invention relates to a variant DNA sequence useful in DNA vaccines.
 CC The DNA sequence encodes a naturally occurring protein such as C3d which,
 CC by virtue of third base redundancy and other variations permissible

CC within an amino acid codon, is non-identical to the naturally occurring
 CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
 CC immunisation vector to encode one or more naturally occurring human or
 CC non-human proteins with immunomodulatory properties. The DNA sequence
 CC is useful for inducing an immune response to an antigen in a human or
 CC animal. A pharmaceutical composition is useful for introducing a DNA
 CC sequence encoding a naturally occurring protein into a human or animal,
 CC by administering the pharmaceutical composition into the human or animal,
 CC where the administration results in a therapeutic effect on the human or
 CC animal. The present sequence is Plasmodium falciparum PfMSP1.19 insert
 CC in pUC105-01 vector

XX Sequence 309 BP; 94 A; 71 C; 69 G; 75 T; 0 other;

Query Match 3.9%; Score 194.6; DB 24; Length 309;
 Best Local Similarity 79.6%; Pred. No. 3e-37;
 Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGACCAATGCTGGAAGAAAGATGCCCCAGAAATAGCGCTGTTTCA 4644
 DB 7 AACATCTCCAGACCAATGCTGGAAGAAAGATGCTGGAAGAAAGATGCTGTTTCA 66
 OY 4645 CATCTGGACGAGCGGCGAAGAGTGCATGCTCTGTAATACAAACAGAGAGATTAAG 4704
 DB 67 CATCTGGACGAGAGAGAGAGATGTAAGTGTCTGTGTAATACAAACAGAGAGATTAAG 126
 OY 4705 TGGGTGGAGAACCCAAACCTACCTGCAATGAAAAAATGGGGGTGTAGCGCGATGCT 4764
 DB 127 TGGGTGGAGAACCCAAACCTACCTGTAACGAGAAACAGTGGATGTAGCGCTG 186
 OY 4765 AAATGACCGAGAGAGAGAGAGCGGCTTAACGAAAGAAATACATGCGAGTGTACTAAG 4824
 DB 187 AAGTGCACCGAGAGAGAGAGCGTGTGTTTACGAAAGAAAGTGTGCTGTAATGTACTAAG 246
 OY 4825 CCCGACTCTATCCACTCTTGCAGCGGATTTTTCGCTCACTTAAT 4873
 DB 247 CCAGACTCTTACCCTTTGTTGATGGAATCTTCTGTTCTCTTAAT 295

RESULT 29
 AAD22460
 ID AAD22460 standard; DNA: 3147 BP.

AC AAD22460;

DT 12-FEB-2002 (first entry)

DE PfMSP1.19-human C3d3 DNA coding sequence from pVK104-01 vector.

KW Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;

KW Immune response; pVK104-01 vector; ds.

XX Chimeric - Homo sapiens;

OS Chimeric - Plasmodium falciparum.

OS Chimeric - Unidentified;

XX Key Location/Qualifiers

FT CDS 1..3147

FT /tag= a

FT /product= "Human C3d3-PfMSP1.19 protein"

FT /note= "CDS does not include stop codon"

FT sig_peptide 1..72

FT mat_peptide 73..3147

FT misc_feature 73..369

FT /tag= d

FT /note= "Antigen sequence"

PN WO200177324-A1.

PD 18-OCT-2001.
XX
XX 09-APR-2001; 2001WO-GB01599.
PF
XX 08-APR-2000; 2000GB-0008582.
PR
XX (ADPR-) ADPROTECH LTD.
PA
XX Steward M, Cox VF;
PI
XX WPI: 2002-010909/01.
DR
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
XX Claim 16; Page 66-68; 87pp; English.
PS
XX The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSPI.19-human
CC C3d3 DNA coding sequence from vector pVK104-01.
XX
SQ Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Query Match 3.9%; Score 194.6; DB 24; Length 3147;
Best Local Similarity 79.6%; Pred. No. 9.9e-37;
Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGCACCATGCGTGAAGAAAGACGCCCCCAAGTAATGCGGCTTTTCACG 4644
DB 73 AACATGCCCCAACCAACGCTTAAGAGCAATGCCAAATCCCGAGTTTTCACA 132
OY 4645 CATCTGAGACGAGCGGAGAGTGAAGTGTCTCTGTAATACAAACAGAGAGATAG 4704
DB 133 CATCTGAGACGAGAGAGAGATGAAGTGTCTCTGTAATACAAACAGAGAGATAG 192
OY 4705 TGCCTGAGAACCCCAACCTTACCTGCAATGAAGAAATGCGGGGTGTGACCGCATGCT 4764
DB 193 TGTGTTGAGAAACCCCAACCTTACCTGTAAGAACACGCGTGAATGCGATGACGCT 252
OY 4765 AATGACACGAGAGAGAGAGCGGCTCTACGGAAGAAATACATGCAAGTGTACTAG 4824
DB 253 AAGTGCACCGAGAGAGAGACTGTGTTCTAACGGAAGAAATTAATTAATGTAAGTACTAG 312
OY 4825 CCGGACTCCTATCACTCTTTCAGCGGATTTTTCCTCCAGCTCTAAT 4873
DB 313 CCAGACTCTTACCTTTTGTTCATGGAATCTTCTTCTCTCTACT 361

RESULT 30
AAD22461
ID AAD22461 standard; DNA; 3147 BP.
XX
AC AAD22461;
XX
DT 12-FEB-2002 (first entry)
XX
DE Human C3d3 DNA-pfMSPI.19 coding sequence from pVK104-02 vector.
XX
KM Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
KW Immune response; pVK104-02 vector; ds.
XX

OS Chimeric - Homo sapiens.
OS Chimeric - Plasmodium falciparum.
OS Chimeric - Unidentified.
XX
XX Key Location/Qualifiers
FT CDS 1..3147
FT /tag= a
FT /product= "Human C3d3-pfMSPI.19 protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 1..72
FT /tag= b
FT mat_peptide 73..3147
FT /tag= c
FT /product= "Mature human C3d3-pfMSPI.19 protein"
FT 2845..3142
FT /tag= d
FT /note= "Antigen sequence"
XX
XX WO200177324-A1.
XX
XX 18-OCT-2001.
PD
XX 09-APR-2001; 2001WO-GB01599.
PF
XX 08-APR-2000; 2000GB-0008582.
PR
XX (ADPR-) ADPROTECH LTD.
PA
XX Steward M, Cox VF;
PI
XX WPI: 2002-010909/01.
DR
XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein -
XX
XX Claim 16; Page 69-71; 87pp; English.

XX The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSPI.19-human
CC C3d3 DNA coding sequence from vector pVK104-02.
XX
SQ Sequence 3147 BP; 869 A; 754 C; 834 G; 690 T; 0 other;

Query Match 3.9%; Score 194.6; DB 24; Length 3147;
Best Local Similarity 79.6%; Pred. No. 9.9e-37;
Matches 230; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 4585 AACATCTCCAGCACCATGCGTGAAGAAAGACGCCCCCAAGTAATGCGGCTTTTCACG 4644
DB 2845 AACATGCCCCAACCAACGCTTAAGAGCAATGCCAAATCCCGAGTTTTCACA 2904
OY 4645 CATCTGAGACGAGCGGAGAGTGAAGTGTCTCTGTAATACAAACAGAGAGATAG 4704
DB 2905 CATCTGAGACGAGAGAGAGATGAAGTGTCTCTGTAATACAAACAGAGAGATAG 2964
OY 4705 TGCCTGAGAACCCCAACCTTACCTGCAATGAAGAAATGCGGGGTGTGACCGCATGCT 4764
DB 2965 TGTGTTGAGAAACCCCAACCTTACCTGTAAGAACACGCGTGAATGCGATGACGCT 3024
OY 4765 AATGACACGAGAGAGAGCGGCTCTACGGAAGAAATACATGCGAGTGTACTAG 4824


```

Db 3025 AACTGCACCGAAGAACTCTGTTCTACGGAAGATTAAGTCCGATCTACTAAG 3084
QY 4825 CCGAGCTCTATCCACTCTGACGGGATTTTTCGCCGCTAAT 4873
Db 3085 CCAGACTCTTACCTTTGTTGATGGAATCTCTGTTCTCTCTACT 3133

RESULT 31
AAD22462
ID AAD22462 standard; DNA; 309 BP.
XX
AC AAD22462;
XX
DT 12-FEB-2002 (first entry)
XX
DE Plasmodium falciparum MSPI.19 (PfMSPI.19) mutant DNA.
XX
KW Immunostimulant; vaccine; immunisation; therapeutic; immune response;
XX MSPI.19 gene; mutant; ds.
XX
OS Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT misc_feature 40..42
FT /tag= a
FT /note= "Mutation occurs at this positions which
FT results in an amino acid change from Cys to Ile"
FT misc_feature 88..90
FT /tag= b
FT /note= "Mutation occurs at this positions which
FT results in an amino acid change from Cys to Trp"
XX
PN WO200177324-A1.
XX
PD 18-OCT-2001.
XX
PF 09-APR-2001; 2001WO-GB01599.
XX
PR 08-APR-2000; 2000GB-0008582.
XX
PA (ADPR-) ADPROTECH LTD.
XX
PI Steward M, Cox VF;
XX
DR WPI; 2002-010909/01.
XX
PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein
XX
PS Example 13; Page 72; 87pp; English.
XX
CC The invention relates to a variant DNA sequence useful in DNA vaccines.
CC The DNA sequence encodes a naturally occurring protein, such as C3d which,
CC by virtue of third base redundancy and other variations permissible
CC within an amino acid codon, is non-identical to the naturally occurring
CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA
CC immunisation vector to encode one or more naturally occurring human or
CC non-human proteins with immunomodulatory properties. The DNA sequence
CC is useful for inducing an immune response to an antigen in a human or
CC animal. A pharmaceutical composition is useful for introducing a DNA
CC sequence encoding a naturally occurring protein into a human or animal,
CC by administering the pharmaceutical composition into the human or animal,
CC where the administration results in a therapeutic effect on the human or
CC animal. The present sequence is Plasmodium falciparum PfMSPI.19 mutant
CC DNA used in the exemplification of the invention.
XX
SQ Sequence 309 BP; 95 A; 71 C; 69 G; 74 T; 0 other;

```

Query Match 3.9%; Score 191.4; DB 24; Length 309;
 Best Local Similarity 78.9%; Pred. No. 1.8e-36;

```

Matches 228; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 4585 AACATCTCCAGCAGCAAAATCGTGAAGAAACAGTCCCGCAGATAGCGGCTTTACAG 4644
Db 7 AACATTGCCCAACACCAACATCCGTTAAGAACAAATTCACAAAACATCCGATTTTCAGA 66
QY 4645 CATCTGGAGACGAGCGGCGAAGAGTGTCTCTGAACTACACAAACAGAGAGATTAAG 4704
Db 67 CATCTGACGAGAGAGAGAGATGGAAGTGTCTGTGAACATACAAAGAGAGATGATTAAG 126
QY 4705 TGGGTGGAGAACCCAAACCTTACCTGCAATGAAGAAATGGCGGTGTGACGCCGATGCT 4764
Db 127 TGTGTTGAGAACCCAAACCTTACCTGTAACGAGAACACGCTGATGCGACGCTAAG 186
QY 4765 AATGCACGAGAGAGACAGACGCTCTAACCGAAGAAATATCATGCGAGTACTAAG 4824
Db 187 AACTGCACCGAAGAAACATCTGTGTTCTAACGGAAGAAATTTCTTCCGATTTACTAAG 246
QY 4825 CCGAGCTCTATCCACTCTGACGAGGATTTTTCGCCGCTAAT 4873
Db 247 CCAGACTCTTACCTTTGTTGATGGAATCTCTGTTCTCTCTACT 295

RESULT 32
AAD22463
ID AAD22463 standard; DNA; 3147 BP.
XX
AC AAD22463;
XX
DT 12-FEB-2002 (first entry)
XX
DE Mutant PfMSPI.19-human C3d3 DNA coding sequence from pVK104-03 vector.
XX
KW Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
KW Immune response; pVK104-03 vector; ds.
XX
OS Chimeric - Homo sapiens.
XX OS Chimeric - Plasmodium falciparum.
XX OS Chimeric - Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..3147
FT /tag= a
FT /product= "Human C3d3-PfMSPI.19 protein"
FT /note= "CDS does not include stop codon"
FT /partial
FT sig_peptide 1..72
FT /tag= b
FT mat_peptide 73..3147
FT /tag= c
FT /product= "Mature human C3d3 protein"
FT misc_feature 73..369
FT /tag= d
FT /note= "Antigen sequence"
XX
PN WO200177324-A1.
XX
PD 18-OCT-2001.
XX
PF 09-APR-2001; 2001WO-GB01599.
XX
PR 08-APR-2000; 2000GB-0008582.
XX
PA (ADPR-) ADPROTECH LTD.
XX
PI Steward M, Cox VF;
XX
DR WPI; 2002-010909/01.
XX
PT Novel variant DNA sequence useful in DNA vaccine, encodes naturally
PT occurring protein and comprises a sequence non-identical to naturally
PT occurring DNA sequence encoding the protein
XX

```


PS Claim 16, Page 73-75; 87pp; English.

XX

CC The invention relates to a variant DNA sequence useful in DNA vaccines.

CC The DNA sequence encodes a naturally occurring protein such as C3d which

CC by virtue of third base redundancy and other variations permissible

CC within an amino acid codon, is non-identical to the naturally occurring

CC DNA sequence encoding that protein. The DNA sequence is useful in a DNA

CC immunisation vector to encode one or more naturally occurring human or

CC non-human proteins with immunomodulatory properties. The DNA sequence

CC is useful for inducing an immune response to an antigen in a human or

CC animal. A pharmaceutical composition is useful for introducing a DNA

CC sequence encoding a naturally occurring protein into a human or animal

CC by administering the pharmaceutical composition into the human or animal

CC where the administration results in a therapeutic effect on the human or

CC animal. The present sequence is Plasmidum falciparum mutant 'PfPfsP.13-human C3d3 DNA coding sequence from vector pVK104-03.'

CC

XX Sequence 3147 BP; 870 A; 754 C; 834 G; 689 T; 0 other;

Query Match	3.9%;	Score 191.4;	DB 24;	Length 3147;
Best Local Similarity	78.9%;	Pred. No. 6e-36;		
Matches 228;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;

QY	4585	AACTTCCTCCGAGCACCAATATCGTGAAGAAACAGTGGCCCCCAATAATAGGGCGCTGTTTCAGG	4644
QY	4586	AACTTCCTCCGAGCACCAATATCGTGAAGAAACAGTGGCCCCCAATAATAGGGCGCTGTTTCAGG	4645
Db	73	AACTTCCTCCGAGCACCAATATCGTGAAGAAACAGTGGCCCCCAATAATAGGGCGCTGTTTCAGG	132
QY	4645	CATCTGGAGCAGCGCGAAGATGCAAGTGTCTCTGAACTACAAACAAGAAAGAGATTAAG	4704
Db	133	CATCTGGAGCAGAGAGAAAGATGGAAGTGTCTGTGTAATCAACAAGAAAGATTAAG	192
QY	4705	TGCGTGGAGAAACCAAAACCTTACTTCGATGAAACAAATAGGGGGGTGTAGCCGATCT	4764
Db	193	TGTGTGAGAAACCAAAACCTTACTTCGATGAAACAAATAGGGGGGTGTAGCCGATCT	252
QY	4765	AAATGACACGAGGAAGACACAGCGCTTAAACGGAAGAAATAACATGCGAGTGTACTAAG	4824
Db	253	AAATGACACGAGGAAGACACAGCGCTTGTGTCTAAACGGAAGAAATACCTTGGGAATGTACTAAG	312
QY	4825	CCGACATCTCATTCACCTTTCACAGGGATTTTTTGTCTCCAGCTCTAATT	4873
Db	313	CCGACATCTTACCTTGTTCGATGGAATCTTCTGTCTTCCCTTAACCT	361

RESULT 33	
AAD22464	
ID	AAD22464 standard; DNA; 3147 BP.
AC	
AA	AAD22464;
DT	12-FEB-2002 (first entry)
DE	Human C3d3-PfMSPI.19 mutant DNA coding sequence from PVX104-04 vector.
XX	
KW	Human; immunostimulant; vaccine; immunisation; therapeutic; C3d;
KM	Immune response; PVX104-04 vector; ds.
XX	
OS	Chimeric - Homo sapiens.
OS	Chimeric - Plasmodium falciparum.
OS	Chimeric - Unidentified.
XX	
Key	Location/Qualifiers
FT	CDS 1..3147
FT	/*tag= a
FT	/product= "Human C3d3-PfMSPI.19 protein"
FT	/note= "CDS does not include stop codon"
FT	/partial
FT	1..72
FT	sig-peptide
FT	/*tag= b
FT	73..3147
FT	mat-peptide
FT	/*tag= c
FT	/product= "Mature human C3d3-PfMSPI.19 protein"
XX	

PM WO200177324-A1.
 PD 18-OCT-2001.
 XX
 XX
 XX 09-APR-2001; 2001WO-GB01599.
 PF
 XX
 XX 08-APR-2000; 2000GB-0008582.
 PR
 XX
 XX (ADPR-) ADPROTECH LTD.
 PA
 XX
 XX Steward M, Cox VF;
 PI
 XX
 XX DR WPI; 2002-010909/01.
 XX
 XX Novel variant DNA sequence useful in DNA vaccine, encodes naturally
 PT occurring protein and comprises a sequence non-identical to naturally
 PF occurring DNA sequence encoding the protein -
 XX
 XX Claim 16; Page 76-78; 87pp; English.
 PS

Query Match	3.9%;	Score 191.4;	DB 24;	Length 3147;
Best Local Similarity	78.9%;	Pred. No. 6e-36;		
Matches 228; Conservative	0;	Mismatches 61;	Indels 0;	Gaps

Oy	4585	AAATTCCTCCCGACCAATGGGTGTAACAACAACAGTGGCCCCCAAGATAGCGGCTGTTCACG	4644
Db	2845	AACATGTGCCCAACCCATAGCGTTTAAGACGAATTTCCACAAAACCTCCGGATGTTTCAG	2904
Oy	4645	CATCTGCAGCAGCCCGAAGAGTGCACAGTGTCTCCTGACTACCAACGAAGAAGAGATAAG	4704
Db	2905	CATCTGCAGCAGAGAGAAGATGTAAGTGTCTGTTGAACCTACAACGCGAGAACTGTATAG	2964
Oy	4705	TGCGTGGAGAACCCAAACCTACTGCAATGAAAACAATGGCGGGGTGACGCCGATGCT	4764
Db	2965	TGTGTGGAAACCCAAACCTACTGTTAAACGAGAACACGGTGGATGCGCAGCTGACGCT	3024
Oy	4765	AAATGCACCCGAGAGACAGCGGCTTAAACGGAAAAGAAATTCACATGCGAGTCTACTAAG	4824
Db	3025	AAGTGCACCGAAGAAAGACTCTGTGTTCTAACGGAAGAAAGATTTACTTCCGATGTACTAAG	3084
Oy	4825	CCCGACTCTATCCACTCTTGACAGGGATTTTTTGCTCCAGCTCAATT	4873
Db	3085	CCAGACTTTCACCTTTGTTGTAAGTAATCTCTGTTCTTCTCTACT	3133

RESULT 34
AAN50354
ID AAN50354 standard; DNA: 306 BP.
XX
AC AAN50354;
XX
DT 22-OCT-1991 (first entry)
XX
DE Storage-specific, late schizont merozoite malaria antigen insert of
DE plasmid p31-1.
XX


```
KW Malaria; vaccine; plasmid p31-1; ss.
XX Plasmodium spp.
OS
XX
XX
FH Key Location/Qualifiers
FT mat_peptide 3..306
FT /tag= a
FT /label= malaria antigen
XX
XX
PN W08503725-A.
XX
XX
PD 29-AUG-1985.
XX
XX
PF 20-FEB-1985; 85WO-GB00072.
XX
XX
PR 20-FEB-1984; 84GB-0004378.
XX
XX
PA (BIOJ ) BIOGEN NV.
XX
XX
PI Mach B, Perrin L, McGarvey M, Cheung A, Shaw A;
XX
XX
DR WPI: 1985-223371/36.
XX
DR P-PSDB; AAP50303.
XX
XX
PT Prod. of antigens of Plasmodium species, esp. of falciparum - by
PT recombinant DNA methods giving polypeptide(s) for protecting
PT against malaria or for diagnosis.
XX
XX
PS Disclosure: Fig 5; 49pp; English.
XX
XX
CC The DNA sequence encodes a Plasmodium falciparum, Plasmodium vivax,
CC Plasmodium malariae and Plasmodium ovale antigen, which may be used
CC in the diagnosis of malaria and as a vaccine against malaria.
CC Nucleotides 293-422 and 429-530 are used in the preparation of
CC monoclonal antibodies against the malaria antigen.
XX
XX
SQ Sequence 306 BP; 87 A; 49 C; 82 G; 88 T; 0 other:
Query Match 2.4%; Score 119.2; DB 6; Length 306;
Best Local Similarity 70.4%; Pred. No. 6.8e-19;
Matches 159; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 82 TATCAGAGCGTGGTTAAGAACTGGAAGCTTGAAGATGCCCTTACCGATACAGC 141
DB 18 TATCAAGAACTGTCTCAAAAACTAGAGAGCTTTAGAGATCCTATTGACAGGTTATAGT 77
QY 142 CTGTTCCAGAGAAGAGATGCTGATGAAGGACGACGCGCGCTTACAAAC 201
DB 78 TTATTTCAAAAGAAAAAATGATTTAAATGAAGAAACAAGTGAACAGCTGTACAACT 137
QY 202 AGCACACCCGGTTCTAAAGGCTGTGGCTAGCGGTGGCTCCGGTGGCTGTGGCCTCT 261
DB 138 AGTACACCTGTGTTGAGTGTGTTCACTTCAAGTGTGTTCAAGTGTGTTCAAGTGTGTTCA 197
QY 262 GGGGGTTCCGTCCGCTCCGCGGACGCGTGCATCAGTGGCTCAG 307
DB 198 GTTGCTTCAAGTGTGTTCAAGTGTGTTCAAGTGTGTTCAAGTGTGTTCAAGTGTGTTCAAG 243
RESULT 35
AAT05868
ID AAT05868 standard; DNA; 3399 BP.
XX
XX
AC AAT05868;
XX
XX
DT 14-AUG-1996 (first entry)
XX
XX
DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
XX
XX
KW Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
KW immunisation; vaccination; ss.
XX
OS Chicken leucocytozoan.
```

```
XX
XX
FH Key Location/Qualifiers
FT CDS 1..3399
FT /tag= a
FT misc-feature 1150..3218
FT /tag= b
FT /note= "Fragment referred to in the claims, for
FT use as insert in a recombinant vaccine
FT against chicken leucocytozoan disease"
XX
XX
PN JP07284392-A.
XX
XX
PD 31-OCT-1995.
XX
XX
PF 19-APR-1994; 94JP-0080643.
XX
XX
PR 19-APR-1994; 94JP-0080643.
XX
XX
PA (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
XX
XX
PA (KITA ) KITASATO KENKYUSHO SH.
XX
XX
DR WPI: 1996-006311/01.
XX
DR P-PSDB; AAR97866.
XX
XX
PT Chicken leucocytozoan immunogenic protein - used in a recombinant
PT vaccine against chicken leucocytozoan disease
XX
XX
PS Claim 6: Page 6-9; 35pp; Japanese.
XX
XX
CC AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
CC or a fragment of it can be used in a recombinant vaccine to immunise
CC against chicken leucocytozoan disease. The DNA is used in a vector
CC and operatively linked to an expression regulatory sequence as in
CC standard practice.
XX
XX
SQ Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other:
Query Match 2.8%; Score 105.6; DB 17; Length 3399;
Best Local Similarity 43.8%; Pred. No. 4.8e-15;
Matches 563; Conservative 0; Mismatches 709; Indels 12; Gaps 2;
QY 702 CAAGGACATGTGGGAAAGTGAAGATTTATTTAAAGATAAGAACCATCGAA 761
DB 1798 CATGAGAGAGAAAAAGAGAGATGACATGAGAGAAAAAGAGAGCATGAAAGAA 1857
QY 762 CATTACGAGCTGATCGAAGAAATCCAAAAAGACCATAGCAAAATTAAGATCAACCAA 821
DB 1858 GTATATACATGAAGAGAAAAAGAGAAAGTAACATGAGAGAAATTAAGAAAGACAT 1917
QY 822 GGAAGAGAAAAAGAGAGTTGTACAGGCCCGCTAGCAGCTGTCTCTATTAACAACA 881
DB 1918 GAAAGAGTAAATACATGAGAGAAAAAGAGAGTAAACATAGAGAAATTAAGAAAGAA 1977
QY 882 GCTTGAAGAGCCCATTAACCTCATACGCTACGTAGAGAGAGCGCATAGACACCTCAAGAA 941
DB 1978 GACCATGAGAGAGTAAATACATGAGAGAAAAAGAGAGAAATTAAGAAATTAAGAA 2037
QY 942 GAATGAGAAATATCAAGAACTGCTCGACAGATTAAATGAATTAAGAAATTCCTCCGACG 1001
DB 2038 AAGAGAGAGCATGAGAGAGTAAATACATGAGAGAAAAAGAGAGAAATTAAGAAAGAA 2097
QY 1002 CAACCTTGGAACACCCCTTAACACGCTGTGACAGAGAAACAAAGATAGAGACACGA 1061
DB 2098 ATAGAAAAAGAGAGAGCATGAGAGAGTAAATACATGAGAGAAAAAGAGAGAAATTAAG 2157
QY 1062 GAAAGAGATCAAGAGATGCCCAAAACCATTAAGTTCAACATGATTCTCTTTACTGA 1121
DB 2158 GAGAGAAAAAGAGAGAGAGCATGAGAGAGTAAATACATGAGAGAAAAAGAGAGAGTAA 2217
QY 1122 TCCCTTGAGCTGAGTACTTGTGAGAGAGAGAAATTAAGAAATTAAGACATCTCCGCCAA 1181
DB 2218 ACACATGAGAGAGAGAGAGAGAGTAAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGTAAATACAT 2277
```


Oy	1182	AGTGAAGCAAGAAATCAACCGCAACCTTAATGAAATTCCTCCATGGTGTGAGCTACCTCT	1241
Db	2278	GAAGAGAAAAAGAGAAAGTAATATACATGGAAGAGAAAAAGAA-----GAATTAACACAT	2331
Oy	1242	GCTCTTAATAGCATATCAACACCGCTCTCAACGAGCTCAATACCTTCGGTACCTTGATTAA	1301
Db	2332	GAAGAGAAAAAGAGAAAGTAATACATGAAAGAAAGAAAAAGAAAGTAATACATGAAAGA	2391
Oy	1302	CCCCCTTGATTATACGAAAGAACCCCTCTAAGATATCTACACAGACATGAGAGAAAGAA	1361
Db	2392	GA-----AAAGAGAGAGTAATACATGAAAGAAAGAAAAAGAAAGTAATACATGAAAGA	2445
Oy	1362	GTTTATCAACGAATCAAGGAGAGATCAAAATTTGAGAGAGAAATTTGAGAGTGAAGA	1421
Db	2446	GAAGAAAAAGTAACACATGGAAGAGAAAAAGAAAGTAATACATGAAAGAGAAAAAGAA	2505
Oy	1422	GAAGAATTACGAGAACCCGACAGCAAAAAGCTTAAACGATATCACTAAAGATGAAAAAGCT	1481
Db	2506	GAATTAACACATGAGAGAGAAAGAAAAAGTAACATGAGAGAGAAAGAAAAAGTAACAT	2565
Oy	1482	GCTGAACGAGATCTATGATTCCAAATTTCAACATTAACATCGACCTGACCACTTGAGAA	1541
Db	2566	GAAGAGAGAGAAAAAGTAACATATGGAAGAGAAAGAAAGAAAGAAAGAAAGTAACAT	2625
Oy	1542	AATGATGGAAAAACGCTACTCTTACAAAGGTGAGAACTGCACACCACTAATATCCTTTC	1601
Db	2626	GAAGAGAGAGAAAAAGTAACATGACATGGAAGAGAAAGAAAGAAAGTAACATGAAAGAGAA	2685
Oy	1602	ATCTTATGAGAAATTTAAGCATATCTTGAGAAAGCTCACCAAAAGCTTTAAGTATATGA	1661
Db	2686	AAATTAATACATGAGAGAGAAAGAAAAAGAAAGATGAGAGAGAAAGAAAGAAAGAA	2745
Oy	1662	GGACTATTCCTCGGGAACATTTGTTGTGGAGAAAGAAACATAAGTATTAACAGAAATCTAT	1721
Db	2746	GAAGAGAGAGAGAGAGAGAGAAAGAAAGTGAAGAGAGAGAAAGAAAGAAAGAAAGAT	2805
Oy	1722	AAGTAAGATCGAAAAACGAGATCGAGACGGCTTGTGGAACATTAAAGAGAGATGAAGAA	1781
Db	2806	GAGAGAGAGAGAGAGAGAGAAATTAAGAGAAAGAAAGAAAGAAAGAAATTAAGAGAA	2865
Oy	1782	GTTGTTGAGAGAGAGATTCAAAAGACGAAATTAACACGATGAGAGAGATCTGGAGGT	1841
Db	2866	GAAGAGAGAGAGAGAGAGAGCATGGAAGAGAAAGTAAACATGAAGAGAGAGAGAGAA	2925
Oy	1842	CTCCGATATTGTTAAAGTCCAAAGTCGAGAGAGGTGCTTCATGAACCAAGATTGAATGAAC	1901
Db	2926	GTAACACATGAGAGAGAGAGAAAGTAACATGAAGAGAGAGAAAGTAATGAATGAATGA	2985
Oy	1902	CAGAGAGACTCAACTCTTCTGGAAGACGTGAGTTTAAAAACATTAATATATATGTCGCCAA	1961
Db	2986	GAAGAGAGAGAGAGAGAGTAACATGGAAGAGAGAGAGAGAGAGAGTAACATGAAGAGAA	3045
Oy	1962	TAGTTATTAAGCAGAGCAATTAAGCA	1985
Db	3046	AAAGTAACACATGAGAGAGAGAA	3069

XX	RESULT 36
XX	AA500655
ID	AA500655 standard; DNA; 618 BP.
XX	
AC	AA500655;
XX	
DT	07-SEP-2001 (first entry)
XX	
DE	Plasmodium vivax merozoite surface protein C-terminal region DNA.
XX	
KW	Merozoite surface protein; malaria; blood; serum; diagnosis; vaccine; ds
KW	antigen-antibody composite; Enzyme Linked Immunosorbent Assay; ELISA;
XX	MSP.
XX	
OS	Plasmodium vivax.
XX	

FH	Key	Location/Qualifiers
FT	CDS	1..618
FT		/*tag= a
FT		/partial
FT		/product= "P. vivax merozoite surface protein
FT		C-terminal region"
FT		/note= "No start or stop codon"
PN		WO200136587-A2.
PD		25-MAY-2001.
XX		
PF		15-NOV-2000; 2000MO-KR01302.
XX		
PR		15-NOV-1999; 99KR-0050616.
PR		25-APR-2000; 2000KR-0022041.
XX		20-MAY-2000; 2000KR-0027305.
PA		(HUMA-) HUMAMBIO CO LTD.
XX		
P1		Park H;
XX		
DR		WPI: 2001-343809/36.
DR		P-PSDB; AAU00669.
PT		New gene encoding merozoite surface protein of Plasmodium vivax, useful
PT		for producing protein for diagnosis of malaria and for vaccination -
XX		
PS		Claim 5; Page 23; 24pp; English.
XX		
CC		The sequence represents a DNA encoding the Plasmodium vivax merozoite
CC		surface protein (MSP) C-terminal region. The C-terminal region of the
CC		merozoite surface protein has a strong antigenicity in malarial diseases.
CC		For diagnosis of malaria, recombinant proteins with enhanced
CC		antigenicity, obtained by addition of fusion proteins to surface protein
CC		C-terminal regions, can be reacted with serum or blood of a Plasmodium
CC		infected patient. Antigen antibody composites will be formed, and these
CC		are detected by Enzyme Linked Immunosorbent Assay (ELISA). The
CC		recombinant antigens provide a quick and reliable diagnosis of malaria,
CC		with good sensitivity and selectivity.
XX		
SO		Sequence 618 BP; 207 A; 114 C; 147 G; 150 T; 0 other;
	Query Match	2.1%; Score 104; DB 22; Length 618;
	Best Local Similarity	54.9%; Pred. No. 4.9e-15;
	Matches 234; Conservative	0; Mismatches 180; Indels 12; Gaps 1
OY	4488	TCTCCTGACCAAGTTTCTGTCCACTGGCATGGTGTCGAAAACCTCGCCAAACAGTGGT 4547
Dd	189	TCTTGTGAAAAATGATGTAATAAATGATTAAAGAAAGCAAGCTCCAAGGAATATT 248
OY	4548	GAGCAATCTGCCTGCAGCGCACCTGCAGGGCATGCGAACATCTCCAGCACCATTGGCT 4607
Dd	249	ATCCCAAGCTGCTAAATTGTCACAACCTCAGTTATTAACTATGAGCTCCGACACACATGTAT 308
OY	4608	GAGAAGACAGTGGCCCCAGAAATAGCGGCTTTTCAGGCCATCTGACGAGCGGAAGAGTG 4667
Dd	309	AGACACCAATGTCCTGATTAATGTCAGCGCTCGTATAGTGACGGAATGAAAGAATG 368
OY	4668	CAAGTGTCTCTGAACTACAAACAAGAGAGATTAAGTSCGTGGAGAACCCCAACCTTAG 4727
Dd	369	GAGATGCTGTGTTAACTTTAAAGAAAGGCGGCAAGTGTGGCCAGATTCAAATGTGAC 428
OY	4728	CTGCATGAAAAATGAGGGGGGTGGAGCGCCATGCTAAATGACACGAGGAAGACAGCG 4787
Dd	429	TGTGTAAGGTAAACAATGGTGGTGTGGCCCTTAAGCTGATATTAATTAATGACGACAGC-- 486
OY	4788	CTCTAACGAAAAAATACATGCGAGTGTACTAAGCCGCACTCTTACACTCTTGA 4847
Dd	487	-----AATAAATGCTGTGTAATGTGCAAAAAGAAAGTGTGAGCCACTTTGA 536
OY	4848	CGGATTTTTTGTGCTCCAGCTTAATTCTCGGGCATTCCTTCTGCTGATCTCATGGCT 4907

Db	537	GGGCGTTTCGTAAGCTCCACGCTCCTAAGCTGTGCTTCTGTATGGCTGATGTTGCT	5368
OY	4908	GATCCT	4913
Db	597	TTTCCCT	602

```

RESULT 37
AAT93729
ID AAT93729 standard; DNA; 165 BP.

```

AC	AAT93729;
XX	
DT	03-APR-1998 (first entry)

DE DNA encoding signal peptide 2 which is used in a malaria vaccine

KM Signal peptide; malaria; vaccine; merozoite surface antigen-1 peptide;
KM MSA-1; recombinant vaccinia virus; Plasmodium falciparum;
KM Immune response; humoral; cell-mediated; merozoite; ss.

OS Mammalia.

PN WO9726911-A1.

PD 31-JUL-1997

29-JAN-1997; 97WO-US01395

PR 29-JAN-1996; 96US-0593006.

PA (GEOU) UNIV GEORGETOWN.

Davidson EA, Yang S,

DR WPI; 1997-393372/36.

DR P-PSDB; AAW34596.

PT Malaria vaccine - comprises expression vector expressing fragment of
 PT merozoite surface antigen

PS Disclosure; Page 19; 75pp; English.

The present sequence encodes a signal peptide that is used in a malaria vaccine. A novel expression vector contains a sequence encoding an immunogenic merozoite surface antigen-1 peptide (MSA-1), and a mammalian signal (AMW34595-96) and/or anchoring sequence (AMW34598). The signal peptide contains 3 regions, a first or c region at the carboxy end of the peptide which serves as the cleavage site for a signal peptidase enzyme, a second or h region which is N-terminal to the c region (and highly hydrophobic), and a third region or n region. The vectors (or analogues that express MSA-1 without either signal or anchor peptides), particularly in the form of recombinant vaccinia virus, are used in vaccines to prevent or treat malaria caused by *Plasmodium falciparum*. The chimeric proteins can also be used in vaccines. The vaccinia vector expresses the antigen fragment for many days, or even years, generating a long-lasting immune response (humoral and/or cell-mediated) against the merozoite form of the parasite, in humans or other animals.

Sequence 165 BP; 65 A; 16 C; 24 G; 60 T; 0 other;

Query Match	2.0%	Score 99.4	DB 18	Length 165
Best Local Similarity	75.2%	Pred. No. 3.3e-14		
Matches 124; Conservative	0	Mismatches 41	Indels 0	Gaps 0

Oy	10	ATGAAATCATTTTCTTCCTCTGTCATTTCTGTTTTATCATCATTA	CTAGTGGCTG	69
Db	1	ATGAAGATCATATTTCTTTTATGTTTCATTTCTTTTTTATTAATAA	TACACATATG	60
Oy	70	ACCACGAACTCCTATCAGAGCTGTTAAGAACTGAGAGCTTGGAAGA	TATGCCGCTCTT	129
Db	61	ACACTGTAAAGTTTATCAGAGCTCTGTCAAAAATCTAAGAGCTTGA	AGAGTGCAGTTTG	120

Qy 130 ACCGATACAGCCTGTTCAGGAAGAGAGATGGTCTGAATGAA 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACAGGTTATAGTTTATTTCAAAAGGAAAAAATGGTATTTAATGAA 165

RESULT 38	
AAI70929	
ID AAI70929 standard; DNA; 16S BP	

AC	AAI70929;
XX	
DT	12-MAR-2002 (first entry)

DE DNA encoding signal peptide.

KW Merozoite surface antigen; 1; MSP-1; p15MSP-1; antigen; immunogen;
KW malaria; vaccine; Venezuela equine encephalitis virus; DVEE;
KW vector; immunisation; Plasmodium falciparum; signal peptide; ss.

Synthetic

PN W0200185927-A1

PD 15-NOV-2001.

08-MAY-2001; 2001WO-US14716

PR 08-MAY-2000; 2000US-202430P

PA (GEOU) UNIV GEORGETOWN

PI Davidson E, Nikodem D;

DR WPI; 2002-049444/06

DR P-PSDB; AAM50529.

Novel vaccine for immunizing mammals against *Plasmodium falciparum* PT
infection, comprises a viral vector system expressing protein PT
corresponding to specific domain of major merozoite surface protein I
of *Plasmodium falciparum*, - PT

PS Disclosure; Page 20; 60pp; English.

The present sequence is that of DNA encoding a typical mammalian signal peptide sequence (see AAM50529). Such a signal peptide sequence may be incorporated into chimeric proteins of the invention that also include the Plasmodium falciparum erythrocytic stage major merozoite surface antigen (MSP-1) immunogenic peptide, p15MSP-1 (see AAM50527). The invention relates to a malaria vaccine comprising an expression vector, preferably a defective Venezuelan equine encephalitis (VEE) viral vector system, which expresses p15MSP-1 or its immunogenic fragment after administration to a patient. The p15MSP-1 protein or fragment raises a humoral and/or cell-mediated response to the erythrocytic merozoite malaria antigen, protecting the patient from a subsequent malaria infection. The VEE viral vector system continues to express antigen in the CC patient for a period of days, months or even years. Inclusion of a signal peptide and/or an anchor peptide sequence in the p15MSP-1 antigen produces an immunogenic response which is significantly (i.e. at least about 2 times and as much as 100 times or more) greater than the immunogenic response produced by p15MSP-1 which does not contain the signal or anchor peptide sequence.

SQ Sequence 165 BP; 65 A; 16 C; 24 G; 60 T; 0 other;

Query Match	2.0%	Score 99.4	DB 24	Length 165
Best Local Similarity	75.2%	Pred. No. 3.3e-14		
Matches 124	Conservative	0	Mismatches 41	Indels 0
				Gaps 0

QY 10 ATGAATAATCATTTCTTCCTCGTGCATTCTGTTTTTATACAACTACTCAGTGCGTG 69
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGAGATCAATATCTTTTATATGTTCATATCTTTTATTAATTAATAAATCACACAATGTGTA 600

PR 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX MPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 31099; 21bp + Sequence listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB161840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 7588 BP; 1405 A; 1605 C; 2192 G; 2386 T; 0 other;

Query Match 1.7%; Score 85; DB 23; Length 7588;
 Best Local Similarity 44.7%; Pred. No. 7.4e-10;
 Matches 377; Conservative 0; Mismatches 460; Indels 6; Gaps 1;

QY 1247 ATATGATATCAACACGCTCTCTCAAGAGCTTCATAGCTTCGCTGACTTGTATTAACCCCT 1306
 DB 1735 ATGAGAACACACACCAATATACAGAGACAAACATCCAAAGCAAAATGAGACAAACC 1676
 QY 1307 TCGATATATACGAAAGAACCCCTCTAAGATATCTACACAGCAATGAGAGAAAGATTTA 1366
 DB 1675 AAATATACATCCAAAGCGAGATGAGAACAAACAAATTAACATCCAAAGCAAGATC 1616
 QY 1367 TCAACGAATCAAGAGAGAGATCAAAATTTGAGAGAAAGAAATTTGAGAGACAGAAAA 1426
 DB 1615 AGAACCAACCAAGCAAGACATCCAAAGTATGATGAGAACAGCAACCAACCAACG 1556
 QY 1427 GTTACGAAGACCGAGCAAAAGTCTAAGCATATCACTAAAGATGAAAGCTGTGTA 1486
 DB 1555 AGAACCAACATCCAAAGCAAGAGTCAAAACATCGAAAGCAAGATCAGAACACCA 1496
 QY 1487 AGAGATCTATGATTCCTCAAAATTCACAAATACATGACCTGACCACTTGGAGAAATGA 1546
 DB 1495 ACAATCCCAAGCAAGATGAGAACCAACCAAGCAAGCAAGATCCAAAGCAAGATGAG 1436
 QY 1547 TGGGAAACCGTACTCTTCAAAAGTGAAGAACTGACACACCATATATCTTGCATCT 1606
 DB 1435 ACAACCAACCAAGCAAGATCCAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1376
 QY 1607 ATGAGATTTTAAAGCATATCTTGAAGAGCTCCAAAGCTTAAATATATGAGAGACT 1666
 DB 1375 GCAAGACTCAAAACACACAGACATCTCCAAAGCAAGATGAGACAAAGCAAGCAACT 1316
 QY 1667 ATTCTCTGCGGAACATTGTTGTGAGAGAAAGCAATTAAGTATTAACAAATCTCATAGTA 1726
 DB 1315 -----TCCAAGCAAGATCAGAACAAACATCCAAAGCAAGATCAGAACCAAGTACA 1262
 QY 1727 AGATCGAAAACGAGATCGAAGCGCTTGTGAGAACATTTAAGAGAGTGAAGAACAGTTGT 1786
 DB 1261 ACATCCCAAGCAAGATCAGAACCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1202
 QY 1787 TTGAGAGAGATTTACAAAGAGCAAGAAATTAACAGATGAGAGATCTCTGAGAGTCCGC 1846
 DB 1201 ACAACCAAGCAAGATCCAAAGCAAGATGAGAACCAAGCAAGCAAGCAAGCAAGCA 1142

QY 1847 ATATTTTAAATCCAGATGAGAGAGTCTCTCATGAAACAGATGATGAACTCAGACA 1906
 DB 1141 ACCAGATTAACATCCAAAGCAAGAAATCAAGAACCAACCAAGAAATTAACATCCAAAGCA 1082
 QY 1907 AGACTCAACTCATTTCTAAGAGCTGGAGTTTAAACATATATATCATGTGCCGAATGTT 1966
 DB 1081 ATCAGAACCAACCAAGCAAGATTAACATCCAAAGCAAGATGAGAACCAACCAAGCAATACA 1022
 QY 1967 ATAAAGAGAGATTAACAGAACCAACCAAGATTAACATCTCATGATGAGAAAGAGATAGACA 2026
 DB 1021 TCCAAGCAAGATCAGAACCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 962
 QY 2027 AACTGAAAGTTCATCCGCAAGAGCTGATCAACGAGAGAGAGAGAAATTA 2086
 DB 961 AAAGCAAGATCAGAACCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAATCC 902
 QY 2087 AAA 2089
 DB 901 AAA 899

RESULT 41

AA087587
 ID AA087587 standard; DNA; 1686 BP.

AA087587;
 AC
 XX
 XX

DT 19-DEC-1995 (first entry)
 DE
 XX

DE DNA encoding Leucocytozoan protozoa structural protein epitope.
 XX
 XX

XX Leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
 KW Leucocytozoanosis; treatment; ss.
 XX

XX Leucocytozoan protozoa sp.
 OS
 XX

PN JP07089995-A.
 PD
 XX

PD 04-APR-1995.
 PF 10-SEP-1993; 93JP-0226078.
 PR 10-SEP-1993; 93JP-0226078.
 XX

XX (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
 PA (NISS-) NISSEIKEN KK.
 XX

XX WPI; 1995-167252/22.
 DR P-SDB; AAR70491.
 XX

XX Immune inducing polypeptide against Leucocytozoan protozoa - useful
 PT in production of vaccines for treatment of leucocytozoanosis in
 PT fowl.
 XX

PS Claim 1; Page 12-14; 20pp; Japanese.
 XX

CC AA087587-89 encode polypeptides having a whole or partial epitope of a
 CC structural protein of Leucocytozoan protozoa (see AAR70491-93). The
 CC polypeptides and DNA encoding them are useful in the production of
 CC vaccines for the treatment of leucocytozoanosis of fowl.
 XX

XX Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;
 SQ

Query Match 1.7%; Score 81.6; DB 16; Length 1686;
 Best Local Similarity 44.4%; Pred. No. 2.3e-09;
 Matches 327; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

QY 1354 AGAAGAGATTTCACAGCAAGATCAAGAGATCAAAATTTGAGAGAGAAATTTGAG 1413
 DB 86 AAATGTAG 145
 QY 1414 AGTGAACAG 1473

Db 146 AACAGAACAGAGAACAGAACAGAAATCGTAGAACAGAACAGAACAGATGAAGAAG 205
 QY 1474 GAAAAGCTCTGACAGATCTATGATTCCTCAATTCACAAATACCTCCAGCCGACCAAC 1533
 Db 206 AACAGAACAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265
 QY 1534 TTCCAGAAATATGAG 1593
 Db 266 AAGAGAACAG 325
 QY 1594 ACCCTTGATCTCTGAG 1653
 Db 326 AAGAGAACAG 385
 QY 1654 TATAG 1713
 Db 386 AAGAGAACAG 445
 QY 1714 AATCTCATATGATGAG 1773
 Db 446 AAGAGAACAG 505
 QY 1774 GAG 1833
 Db 506 AAGAGAACAG 565
 QY 1834 CTGAG 1893
 Db 566 ATGCTGAG 625
 QY 1894 GATGAG 1953
 Db 626 AAGAGAACAG 685
 QY 1954 GAG 2013
 Db 686 ATTTTAAAG 745
 QY 2014 AAG 2073
 Db 746 ATTAG 805
 QY 2074 AAGAGAACAG 2089
 Db 806 AAGAGAACAG 821
 RESULT 42
 ABL33103/C
 ID ABL33103 standard; DNA: 7758 BP.
 XX
 AC ABL33103;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1076.
 XX
 KW Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antileukemia; anti-HIV; anticonvulsant; ophthalmological;
 KW antineoplastic; anti-HIV; anticonvulsant; ophthalmological;
 KW antineoplastic; anti-HIV; anticonvulsant; ophthalmological;
 KW antineoplastic; anti-HIV; anticonvulsant; ophthalmological;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene: ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000: 2000DE-1032529.
 PR 01-SEP-2000: 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1: SEQ ID NO 1076; 32bp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 7758 BP; 1608 A; 44 C; 1838 G; 4268 T; 0 other;
 Query Match 1.7%; Score 81.6; DB 24; Length 7758;
 Best Local Similarity 42.6%; Pred. No. 5.1e-09;
 Matches 481; Conservative 0; Mismatches 644; Indels 3; Gaps 1;
 QY 693 CGAGACATCAAG 752
 Db 2263 CTTAAAC 2204
 QY 753 CATGAG 812
 Db 2203 CCAC 2144
 QY 813 TGCAC 872
 Db 2143 CCAC 2084
 QY 873 TAAC 932
 Db 2083 CCAC 2024
 QY 933 CCTCAAG 992
 Db 2023 TCTAAAC 1964
 QY 993 TCCGAC 1052
 Db 1963 CTAAAC 1904
 QY 1053 GGAGAC 1112
 Db 1903 CTTAAAC 1844
 QY 1113 CTTTACTGATCCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172
 Db 1843 CCAC 1784
 QY 1173 CTCCGAC 1232
 Db 1783 CTAAAC 1724
 QY 1233 GTACCTCTGCTTATAG 1292
 Db 1723 CTTAAAC 1664
 QY 1293 CTTGATTAAC---CCCTTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1349

Db 1663 CCAAAACACAAATCTACAACTAAACCACTAAACCAACAAACAAACCAACCAACCAAC 1604
QY 1350 TGAGAGAAAGAGTTTATCAACGAAATCAAGAGAGATCAAAATTTGAGAGAAAT 1409
Db 1603 CTTAAAAAAGAAAAAAGACCTAAACACCTATTAACACAAAAAAGACCTAAACAT 1544
QY 1410 TGAGAGTACAGAAAGATTAAGAAACCCAGCAAGGCTTAACGATTAACCTAAGA 1469
Db 1543 CTATATACACAAAAAAGACCTAAACCAACAAACAAATTAACACCTAAACAA 1484
QY 1470 GTATGAAAGAGCTGAGACGATCTATGATTCGAATTAACATTAACATGACCTGAC 1529
Db 1483 CAATATAAACACTTAAACCTTAAACAAACAAACAAACAAACCTTAAACAA 1424
QY 1530 CAACCTTGAGAAATGATGGAAACGCTCTTACAAAGTGAAGAACTGACACCA 1589
Db 1423 CAAACAAACAAATTAACATCCCAAAACAAATTAACAAACCAACCAACCTTAAACAA 1364
QY 1590 TAATACCTTTGATCTTATGAGAAATTTGAATTAATCTTGAAGAGCTACCAAGCTCT 1649
Db 1363 AAAAAAAGACCTAAACCTCCCAAAACAAACAAACAAACAAATTAACACCTTAAACAA 1304
QY 1650 TAAGTATATGAGAGACTATCTCTGCGGAACATTTGTGTGAGAAAGAACTAAAGTATTA 1709
Db 1303 CAAAAAAGACACTTAAACCTCCCAAAACAAACAAATTAACAAACCAACCTTAAACAA 1244
QY 1710 CAAGAATCTCATAGTAAAGTGAAGAGATGAGAGAGCTTGTGAGAACTTAAGA 1769
Db 1243 CAAAAAAGACCTTAAACCTCCCAAAACAAACAAACAAACAAACCTTAAACAA 1184
QY 1770 GGATGAAAGACGTTGTTGAGAAAGATTAACAAAGACGAAATTA 1817
Db 1183 CAAAAAAGACACTTAAACCTTAAACCAACAAACAAACAAACCTTAAACAA 1136

RESULT 43
AAH47054
ID AAH47054 standard; DNA: 5438 BP.
XX
AC AAH47054;
XX
DT 29-OCT-2001 (first entry)
XX
DE Synthetic gene ViVac encoding recombinant protein ViVac1p.
XX
KW Multivalent protein; Immune response; Plasmodium vivax; parasite;
KW protozoa; vaccine; malaria; recombinant; ViVac1; ds.
XX
OS Synthetic.
OS Plasmodium vivax.
XX
FH Key Location/Qualifiers
FT CDS 7..5430
FT /tag= a
FT /product= "ViVac1p"
XX
XX WO200155181-A2.
PN
PD -02-AUG-2001.
XX
PF 29-JAN-2001; 2001MO-US02937.
XX
PR 31-JAN-2000; 2000US-0179213.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Lai AA, Xiao L, Zhou Z;
XX
XX WPI: 2001-514557/56.
DR P-PSDB; AAB85697.
XX
XX New recombinant multivalent protein comprising antigenic determinants
PT derived from more than one stage in a life cycle of Plasmodium vivax.

PT useful as a vaccine for treating, preventing and reducing malaria
XX infection
XX
PS Claim 9; Page 36-39; 59pp; English.
XX
XX The invention relates to recombinant multivalent proteins (I) that
CC stimulate an immune response to Plasmodium vivax. (1) comprises antigenic
CC determinants, fragments or conservative substitutions, derived from more
CC than one stage in a life cycle of a Plasmodium vivax parasite. (1) is
CC useful as a vaccine for stimulating an immune response, specifically a
CC protective immune response that confers increased resistance to infection
CC by Plasmodium parasites, such as P. vivax. (1) is especially useful in
CC the treatment, prevention and reduction of malarial infection, as
CC research or diagnostic reagents for the detection of Plasmodium species
CC in a biological sample, and for conferring immunity against multiple
CC stages of the malarial parasite. The antibodies produced are useful for
CC the detection or measurement of antigenic epitopes derived from one or
CC more stages in a life cycle of a parasite, particularly P. vivax. The
CC vaccine comprising the recombinant proteins, is cost-effective, health-
CC promoting intervention for controlling, preventing or treating the
CC incidence of malaria. The present sequence represents a synthetic gene
CC ViVac1 encoding the recombinant protein ViVac1p, a multivalent and
CC multistage vaccine against P. vivax.
XX
SQ Sequence 5438 BP; 1861 A; 1038 C; 1389 G; 1150 T; 0 other;
XX
Query Match 1.5%; Score 72.4; DB 22; Length 5438;
Best Local Similarity 55.0%; Pred. No. 7.3e-07;
Matches 142; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 4588 ATCTCCAGACCAATGCGGAGAAAGACAGTGGCCCAAGATAGCGCTGTTACGAT 4647
Db 5164 AGCTCCGACACATGATTAAGACCAATGTGCTGATATGACGCTCTTATAGGTAC 5223
QY 4648 CTGAGAGAGCGCGAGAGCTGCAAGTGTCTCTGACATCAACAAAGAGATTAATGTC 4707
Db 5224 TTGGACGAGAAAGATGAGATGATGCTTTTAAGAAAGAGAGCGCAAGTGT 5283
QY 4708 GTGAGAGAACCAACCTTACCTGCAATGAACAAAGTGGGCTGTGAGCGGATGCTAA 4767
Db 5284 GTGCCACATCGAATGTGATGCTTTTAAGATTAACATGTGTGTGCTGCAAGTGA 5343
QY 4768 TGCACGAGAGAGACAGCGGCTTACGAGAAAGAAATACATGCGATGACTAACCC 4827
Db 5344 TGTAAATGACGACACCAATTAATGCTGTAAATGACTAAAGAGGTTCGACCA 5403
QY 4828 GACTCTATTCACATCTTC 4845
Db 5404 CCTTTGAGGAGTTTTC 5421

RESULT 44
AAH47055
ID AAH47055 standard; DNA: 6101 BP.
XX
AC AAH47055;
XX
DT 29-OCT-2001 (first entry)
XX
DE Synthetic gene ViVac2 encoding recombinant protein ViVac2p.
XX
KW Multivalent protein; Immune response; Plasmodium vivax; parasite;
KW protozoa; vaccine; malaria; recombinant; ViVac2; ds.
XX
OS Synthetic.
OS Plasmodium vivax.
XX
FH Key Location/Qualifiers
FT CDS 7..6093
FT /tag= a
FT /product= "ViVac2p"
XX
XX WO200155181-A2.
PN


```

XX 02-AUG-2001.
PD
XX
XX 29-JAN-2001; 2001WO-US02937.
PF
XX
XX 31-JAN-2000; 2000US-0179213.
PR
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Lal AA, Xiao L, Zhou Z;
PI
XX
XX WPI: 2001-514557/56.
DR
XX
XX P-PSDB; AAB85698.
PT
XX
XX New recombinant multivalent protein comprising antigenic determinants
PT derived from more than one stage in a life cycle of Plasmodium vivax,
PT useful as a vaccine for treating, preventing and reducing malarial
PT infection _
XX
XX
XX Claim 9; Page 45-48; 59pp; English.
PS
XX
XX The invention relates to recombinant multivalent proteins (I) that
XX stimulate an immune response to Plasmodium vivax. (I) comprises antigenic
XX determinants, fragments or conservative substitutions, derived from more
XX than one stage in a life cycle of a Plasmodium vivax parasite. (I) is
XX useful as a vaccine for stimulating an immune response, specifically a
XX protective immune response that confers increased resistance to infection
XX by Plasmodium parasites, such as P. vivax. (I) is especially useful in
XX the treatment, prevention and reduction of malarial infection, as
XX research or diagnostic reagents for the detection of Plasmodium species
XX in a biological sample, and for conferring immunity against multiple
XX stages of the malarial parasite. The antibodies produced are useful for
XX the detection or measurement of antigenic epitopes derived from one or
XX more stages in a life cycle of a parasite, particularly P. vivax. The
XX vaccine comprising the recombinant proteins, is cost-effective, health-
XX promoting intervention for controlling, preventing or treating the
XX incidence of malaria. The present sequence represents a synthetic gene
XX multistage encoding the recombinant protein Vivac2p, a multivalent and
XX multistage vaccine against P. vivax.
XX
XX Sequence 6101 BP: 2077 A; 1168 C; 1534 G; 1322 T; 0 other;
SQ
XX
XX Query Match 1.5%; Score 72.4; DB 22; Length 6101;
XX Best Local Similarity 55.0%; Pred. No. 7.8e-07;
XX Matches 142; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
XX
XX 4588 ATCTCCAGACCAATGCGTGAAGAAACAGTCCCGCCAGATAGCGGCTTTCAGGCAT 4647
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5827 AACTCCGAGCACACATGTAATAGACACCAATGTGCTGATAAATGACAGCGTATAGTAC 5886
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4648 CTGACGACGAGCGGAGAGTGAAGTGTCTCTGAACTACAAACAAGAGAGATAGTGC 4707
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5887 TTGGACGAGCAAGCAAGAAATGAGATGCTTGTTAACTTTAAAGAAAGCGGCAAGTGT 5946
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4708 GTGGAGAAACCAACCTTACTGTCGCAATGAAAAAATGCGGGGTGTGACGCCGATGCTAA 4767
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5947 GTGCCAGCATCGAATGTGACTGTGTAAGGATACAAATGCGTGTGCCCCCTGAAGCTGAA 6006
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4768 TGCACCGAGGAAGACAGACGGCTCTTAACGGAAGAAAGAAATATACATGCGAGTGTACTAAGCCC 4827
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 6007 TGTAAATATGACGACAGCAATATAAATCGTCTTAATGTACTAAAGAAAGTTCCTGAGCA 6066
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 4828 GACTCCTATCCACTCTTC 4845
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 6067 CCGTTTGAAGGAGTTTTC 6084
XX
XX
XX RESULT 45
XX AAA70099
XX ID AAA70099 standard; DNA: 3579 BP.
XX
XX AC AAA70099;
XX

```

DT	07-NOV-2000	(first entry)
DE	Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:232.	
XX		
XX	Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;	
KW	antimalarial; malaria; protozoacide; infection; insecticide; ds.	
OS	Plasmodium falciparum.	
XX		
XX	WO200025728-A2.	
PD	11-MAY-2000.	
XX		
XX	05-NOV-1999; 99WO-US26796.	
PF		
XX	05-NOV-1998; 98US-0107131.	
PR		
XX		
PA	(HOFF/) HOFFMAN S.	
PA	(CARU/) CARUCCI D.	
PA	(GARD/) GARDNER M.	
PA	(VENT/) VENTER J C.	
PI	Hoffman S, Carucci D, Gardner M, Venter JC;	
DR	WPI: 2000-365347/31.	
XX		
XX	Proteins encoded by chromosome 2 of the human malarial parasite,	
PT	Plasmodium falciparum, useful as antimalarial vaccines and in the	
PT	diagnosis of P.falciparum infection -	
PS	Disclosure: Page 457-458; 577bp; English.	
XX		
XX	The present invention describes proteins and their fragments (I) encoded	
CC	by chromosome 2 of the human malarial parasite, Plasmodium falciparum.	
CC	Also described are: (1) nucleotide sequences (II) encoding (I); and (2)	
CC	vaccines against P. falciparum infection comprising (I) or (II).	
CC	(I) and (II) are useful for the development of vaccines against	
CC	P. falciparum infection. (I) and polyclonal antisera or a monoclonal	
CC	antibody raised to immunogens comprising the sequences of (I), are	
CC	useful in the detection of infection with P. falciparum. Furthermore,	
CC	(I) (especially when they are rifins or secreted or membrane proteins)	
CC	can aid the identification of drugs to treat or prevent P. falciparum	
CC	infection, or they can be used to identify drug resistance in	
CC	P. falciparum. Sequencing of the Plasmodium chromosome 2 and the	
CC	subsequent identification of proteins encoded by it will help to expand	
CC	our understanding of parasite biology, a process hampered by the	
CC	complexity of the parasitic lifecycle, and provide new targets for	
CC	vaccine and drug development. Parasite resistance to drugs and mosquito	
CC	resistance to insecticides have led to a resurgence of malaria in many	
CC	parts of the world, and there is a pressing need for vaccines and new	
CC	drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide	
CC	and protein sequences given in the present invention, but which are not	
CC	specifically mentioned within the specification.	
XX		
XX	Sequence 3579 BP; 1904 A; 398 C; 552 G; 725 T; 0 other:	
XX		
Query Match	1.4%; Score 70.6; DB 21; Length 3579;	
Best Local Similarity	47.7%; Pred. 1.6e-06;	
Matches 239; Conservative 0; Mismatches 259; Indels 3; Gaps	1.;	
QY	1610 AGAATTCTAAGCATATCTTGAGAAAGCTCCACCAAGCTCTTAAGTATATGAGGACTATT 1669	
DB	1245 AGAAAAACAAGAGAAGAAACAGAGTGTAGACGAAAAAACAAGAAAAAGCGGAGAAAGAAATT 1304	
QY	1670 CTCCTCGGAACATTGTGTGGGAGAAAGAACTAAAGATTACAAAGAAATCTCAATTAAGTAAG 1729	
DB	1305 AGAAGAAGCAAAAGAAAGAAATTCAGAAAAACAAAAAGAAAGAAATCAGAAAAAGATTAAGAAAGA 1364	
QY	1730 TCGAAACAGAGATCGAGACCTTGTTGAGACAAATTAAAGAAAGATGCAAGAACAGTTGTTG 1789	
DB	1365 ATCAGAAAAAGATTAAAGAAAGATTCAGAAAAAACAAGAAAAAACTGAAAGAAAGATGAAAGA 1424	
QY	1790 AGAAGAAGTTTCAAAAAGACGAA---AATTAACCAAGATGAGAAAGATCTCTGAGGTTCTCCG 1846	


```

XX WO200218632-A2.
PN
XX
XX
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP10074.
PF
XX
XX 01-SEP-2000; 2000DE-1043826.
PR
XX 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
XX Sequence 646 BP; 385 A; 97 C; 46 G; 118 T; 0 other;
SQ

```

```

Query Match 1.4%; Score 70.2; DB 24; Length 646;
Best Local Similarity 46.3%; Pred. No. 8.5e-07;
Matches 268; Conservative 0; Mismatches 308; Indels 3; Gaps 1;

```

```

OY 1258 AACACGCTCTACAGAGCTCAATAGCTTGGTGAATTAAGCCCTTCGATTTATACG 1317
DB 20 AACATACGATTAACGATTAACATTAACGATTAACATTAACGATTAACGATTAACAT 79
OY 1318 AAAGAACCTCTAGATATTTACACAGCAATGAGAGAAAGATTTCACAGCAATC 1377
DB 80 AGCAATTAACATTAACATTAATTAACGATTAACATTAACGATTAACGATTAACGAT 139
OY 1378 AAGGAGAGATCAAAATTGAGAAAGAAATTGAGAGTGCACAGAAAGTTACGAGAC 1437
DB 140 AAAATTAACAAATTAATTAACGATTAACATTAACGATTAACATTAACGATTAACGAT 199
OY 1438 CGCAGCAAAAGTCTAACGATATCTAACGATTAAGATTAAGAAAGCTGCTGAAGATCTAT 1497
DB 200 AACATACGATTAACGATTAACATTAACGATTAACATTAACGATTAACGATTAACAT 259
OY 1498 GATTCCAAAATTCACATTAACATTCGACCTGACCACTTGAGAAAATGATGGAAAACGG 1557
DB 260 AGCAATTAACGATTAACATTAATTAATTAACGATTAACATTAACGATTAACATTAAT 319
OY 1558 TACTCTACAAAGTGAGAAATGACACCACTTAATCTTTGCA---TCTATGGAAT 1614
DB 320 AACGATTAACAAATTAACGATTAACATTAACGATTAACGATTAACGATTAACATTAAT 379

```

```

OY 1615 TCTAACATTAATCTTGAGAGCTCACCAAGCTCTTAAGTATATGAGAGCATTTCTCG 1674
DB 380 AACATCAAAATTAACGATTAACATTAACGATTAACGATTAACGATTAACGATTAACAT 439
OY 1675 CGGACATTTGTTGTGAGAGAAAGAACTAAGTATTACAGAAATCTCTAAGTAGATCGAA 1734
DB 440 AGCAATTAACGATTAACATTAATTAATTAATTAACGATTAACGATTAACGATTAACGAT 499
OY 1735 AACGAGATGACAGCGCTTGTGAGAAACATTAAAGAGATGAAAGAACAGTTGTTGGAAG 1794
DB 500 AACGATTAACAAATTAACGATTAACGATTAACGATTAACGATTAACGATTAACGATTAAC 559
OY 1795 AAGATTCAAAAGACGAAATTAACGATTAACGATTAACGATTAACGATTAACGATTAAC 1833
DB 560 AACATACGATTAACGATTAACATTAACGATTAACGATTAACGATTAACGATTAAC 598

```

```

RESULT 48
ABQ39490/C
ID ABQ39490 standard; DNA; 969 BP.
XX
XX ABQ39490;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26081.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP10074.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI: 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the

```


XX	16-NOV-2000	(first entry)	
XX	Neurospora crassa	gde-1 gene.	
XX	Gene silencing: quelling deficient; qde-1; al-1; ds.		
XX	Neurospora crassa.		
XX	Key	Location/Qualifiers	
XX	CDS	2447..6655	
XX	FT	/tag= a	
XX	FT	/product= qde-1	
XX	PN	WO200050581-A2.	
XX	PD	31-AUG-2000.	
XX	PF	16-FEB-2000; 2000WO-IT00048.	
XX	PR	22-FEB-1999; 99IT-RM00117.	
XX	PA	(UYRO-) UNIV ROMA LA SAPIENZA.	
XX	PI	Macino G, Cogoni C;	
XX	DR	WPI: 2000-57917/54.	
XX	PS	P-PSDB: AAB13956.	
XX	PT	Novel polynucleotide encoding a polypeptide which has a silencing	
XX	PS	activity and comprising a RNA-dependent RNA polymerase domain	
XX	Claim 1;	Page 31-43; 48pp; English.	
CC	The present sequence is the Neurospora crassa qde-1 gene. This gene has		
CC	silencing activity. The qde-1 gene was isolated by mutational analysis		
CC	of an al-1 transgenic strain. This strain had an albino phenotype		
CC	resulting from post-transcriptional silencing of the endogenous al-1		
CC	gene. Reversion of this phenotype indicated a mutation in a silencing		
CC	gene. The silencing gene, qde-1, could then be isolated. Modulation of		
CC	qde-1 expression may be used to inactivate genes and to silence		
CC	suppression of genes.		
XX	Sequence 8045 BP; 2089 A; 2180 C; 1950 G; 1826 T; 0 other;		
Query Match	1.3%	Score 65.2; DB 21; Length 8045;	
Best Local Similarity	50.3%;	Pred. No. 5.1e-05;	
Matches 160; Conservative	0;	Mismatches 158; Indels 0; Gaps 0;	
QY	1238	CTCTGCTTTATACGATATACAAACGGCTCTACAGAGCTCAATAGCTTCGGTGA	1297
DB	7390	CTCTACCTGATCATCGGAGACACAAACACAAACAAACAAACAAACAAACAA	7449
QY	1298	TTAACCCTTCGATTTATACGAAAGAACCTCTTACAGATATCTACACAGCAATGAGAGA	1357
DB	7450	ACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	7509
QY	1358	ACAAAGTTTTCACGAAATCAGGAGGAATCATAAATTTGAGAAAGAAATTTGAGAGTG	1417
DB	7510	ACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	7569
QY	1418	ACAAGAAAAGTTACGAAGACCGACAGAAAGTCTTAAAGATATCTCAATAAGATATGAA	1477
DB	7570	ACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	7629
QY	1478	ACCTGCTGACAGATCTTATGATTTCCAAATTAACATATACATCGACCTGACCAACTTG	1537
DB	7630	ACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA	7689
QY	1538	AGAAAATGATGGAAAC 1555	
DB	7690	AGACAAACAAACAAACAAAC 7707	

[illegible]

OY 1437 CCGAGCAAAAGTCTAAACGATATCATAGAGTATGAAAAAGCTGACAGAGATCTA 1496
 DB 2359 AA 2300
 OY 1497 TGATTCGAATTCACATATACATGACCTGACCAACTTCGAGAAATGATGGGAAACG 1556
 DB 2299 AA 2240
 OY 1557 GTACTCTTCAAGTGGGAAACGTGACACCATATATCTTTGATCTCTATGAGATTTC 1616
 DB 2239 AA 2180
 OY 1617 TAAGCATATCTTGTGAGAGCTCACCAAGCTCTTAATATATGAGAGCTATCTCTCG 1676
 DB 2179 AA 2120
 OY 1677 GAACATTTGTTGGAGAAAGACTAAAGTATACAAAGATCTCAATAGATCGAAAA 1736
 DB 2119 AA 2060
 OY 1737 CGAATCGAGCGCTTGTGAGACATTAGAGAGATGAACAGTGTGAGAGAGA 1796
 DB 2059 AA 2000
 OY 1797 GATTACAAAAGACGAAATTAACGATGAGAGA 1831
 DB 1999 AA 1965

RESULT 53

AAV20700 standard; DNA; 5163 BP.

AAV20700;

17-AUG-1998 (first entry)

Cryptosporidium parvum GP900 antigen open reading frame.

Open reading frame; ORF; antigen: GP900; cryptosporidium; infection;

antibody; prophylaxis; treatment; inhibition; retardation;

detection; diagnosis; human; ds.

Cryptosporidium parvum.

Key CDS location/Qualifiers

1..5163

/*tag= a

/product= GP900 antigen

misc_feature 524..1270

/tag= b

/note= "region containing NINC mutations"

WO9806430-A1.

19-FEB-1998.

11-AUG-1997; 97MO-US14104.

14-AUG-1996; 96US-0700651.

(REGC) UNIV CALIFORNIA.

Gut J, Leech J, Nelson RC, Petersen C;

WPI; 1998-159290/14.

P-PSDB; AAW48299.

Anti-Cryptosporidium antibody - used to develop products for

detection, diagnosis, prophylaxis or treatment of Cryptosporidium

PS Claim 32; Pages 60-62; 89pp; English.
 CC The sequence is that encoding the GP900 antigen which may be used
 CC in the production of anti-Cryptosporidium antibodies. These can be
 CC used for the prophylaxis, treatment, inhibition or retardation of
 CC a Cryptosporidium infection in humans or in animals such as calves.
 CC They can also be used for the detection and diagnosis of related
 CC infections.

Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T; 0 other;

Query Match 1.2%; Score 60.6; DB 19; Length 5163;

Best Local Similarity 44.2%; Pred. No. 0.00053;

Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

OY 3930 CAATGTCAAGTGAAGACATTCGACAGCGCTTATATAGAGAGAAATTCAGAA 3989
 DB 615 CAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 674
 OY 3990 CGCTTGGAGAGCGACTTGATTCCTATTAAGACCTGACCTCTACTACGTTGCA 4049
 DB 675 CAACACACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 734
 OY 4050 GGACCATACAAAGTCTCTCAATTAAGAGAGAGGATTAATTTGCTACTACAACTA 4109
 DB 735 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 794
 OY 4110 TATCAAGACTTCATCGACACCGATATCAATTTGCTATGATGCTGGGATTTACA 4169
 DB 795 CAACCACTACTCAACCAACCACTACTCAACCACTACTCAACCACTACTCAAC 854
 OY 4170 GATCCTGAGCGAAATATCAAGTGTGACCTGACCTCTATTAAGATATCAAGATTA 4229
 DB 855 CCACACTACTACCAACCAACCACTACTACCAACCAACCAACCAACCAACCA 914
 OY 4230 GCAAGGCGAGATGAATAATATATGCTGCCCTCTGATATACATCGAAACCTGTACAG 4289
 DB 915 AGAAACCAACAAACAAACCACTACTCAACCAACCAACCAACCAACCACTACTCAAC 974
 OY 4290 AGTAGACACAAATCGACCTCTTGTGATTTACCTGAGAGGCAAGGCTCTCAACTATAC 4349
 DB 975 CCACACCAACCAACCAACCACTACTCAACCAACCAACCAACCACTACTACTA 1034
 OY 4350 TTACGAGAGAGCAATGTGAGAGTTAAATCAAGAGGCTGAACCTCAAAACATCCA 4409
 DB 1035 CTACACACAAACAAACCACTACTACTACTACTACTACTACTACTACTACTA 1094
 OY 4410 AGACAGCTGCGAGATTTCAAGAAATTAACAATTTGCTGGAAATGCGAGACTGTCTAC 4469
 DB 1095 CAACCAACCAACCAACCAACCACTACTACTACTACTACTACTACTACTACTACTA 1154
 OY 4470 CGATTATACCAACAACATCTCC 4492
 DB 1155 CAACCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1177

RESULT 54

AA61849 standard; DNA; 5163 BP.

AA61849;

28-OCT-2000 (first entry)

ORF encoding a portion of Cryptosporidium parvum NINC isolate GP900.

GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;

competitive inhibition; attachment; invasion; ligand binding; sporozoite;

merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.

Cryptosporidium parvum.

Key location/Qualifiers


```
FT CDS 1669..7182
FT /tag- a
FT /partial
FT /product- "Cryptosporidium parvum NINC isolate GP900"
FT /note- "No start or stop codons given in the
FT specification"
XX
XX US6071518-A.
XX
XX 06-JUN-2000.
XX
XX 12-SEP-1997; 97US-0928361.
XX
XX 13-SEP-1996; 96US-0026062.
XX 01-JUN-1993; 93US-0071880.
XX 29-MAY-1992; 92US-0891301.
XX 03-APR-1995; 95US-0415751.
XX 14-AUG-1996; 96US-0700651.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Petersen C;
XX
XX WPI: 2000-422065/36.
XX P-PSDB: AAB11727.
XX
XX New GP900 protein fragments and fusion proteins of Cryptosporidium
XX parvum, useful for detecting the presence of the parasite, and
XX diagnosing or treating Cryptosporidium infections by competitive
XX inhibition of the function of GP900 -
XX
XX Claim 16; Column 47-52; 59pp; English.
XX
XX The invention relates to the GP900 glycoprotein of the protozoan
XX Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
XX proteins comprising GP900 or fragments. The invention also relates to the
XX administration of GP900 or fragments thereof to a host to elicit anti-
XX GP900 antibody production, and to a method of cryptosporidiosis treatment
XX or prophylaxis comprising administration of anti-GP900 antibodies to an
XX individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
XX competitively inhibit sporozoite or merozoite attachment or invasion, and
XX are also useful for the generation of anti-GP900 antibodies. The
XX antibodies also inhibit sporozoite or merozoite attachment/invasion and
XX additionally inhibit the binding of GP900 ligands to GP900. GP900
XX proteins, fragments and antibodies may therefore be used to treat or
XX prevent cryptosporidiosis. Infection with Cryptosporidium is a common
XX cause of diarrhoea in humans and causes life-threatening diarrhoea in
XX immunocompromised persons. Cryptosporidiosis can be contracted from
XX contaminated municipal water supplies (e.g., public swimming pools). It
XX is also a cause of disease in animals, resulting in financial losses in
XX agriculture. GP900 fragments, fusion proteins and antibodies may also be
XX used for the diagnosis of Cryptosporidium parvum infections, and for the
XX detection of the parasite in the environment. The present sequence
XX represents the open reading frame (ORF) encoding a portion of the GP900
XX protein of the NINC isolate of Cryptosporidium parvum.
XX
XX Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other;
XX
XX Query Match 1.2%; Score 60.6; DB 21; Length 5163;
XX Best Local Similarity 44.2%; Pred. No. 0.00053;
XX Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;
XX
XX 3930 CAATGTCACGTGAGAGACATTCTGAACAGCCGCTTTAATAGAGAAATTTCAAGAA 3989
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 614 CAACACACACACACACACACACACACACACACACACACACACACACACACACAC 673
XX
XX 3990 CGTCTTGAGAGAGGAGTGTATTCCTTAATAGACCTGACCTCTTAATAGCTTGTCAA 4049
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 674 CAACACACACACACACACACACACACACACACACACACACACACACACACACAA 733
XX
XX 4050 GGAGCCATACAACTGCTCTCAATAAGAGAGAGGATTAATTTCTGTCTGTACAACTA 4109
XX ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 734 CGAGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
```

```
QY 4110 TATCAAGAGACTCCATGACAGCCGATATCAATTTGCTAATGATGCTGGGTATTACAA 4169
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 794 CAACACACACTACACACACACACACACACACACACACACACACACACACACAA 853
QY 4170 GATTCGTGAGGAAATATACAGTCTGACCTTATTAATAAGATATATCAACGATTA 4229
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 854 CCACAACTACACACACACACACACACACACACACACACACACACACACACAC 913
QY 4230 GCAGGGGAGATGAATAATATCTGCCCTTCGAAATACATGAAACCTGTACAGAGC 4289
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 914 AGAACCACACACACACACACACACACACACACACACACACACACACACACAC 973
QY 4290 AGTGAAGAGCAAAATGACCTCTTGTAAATTCACCTGAGGCCAGGTCTCTAATATAC 4349
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 974 CCACACACACACACACACACACACACACACACACACACACACACACACACAC 1033
QY 4350 TTACGAGAGAGCAATGTGAGTAAATCAAGAGCTGACTACCTCAAAACAAATCCA 4409
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1034 CTACCCACACACACACACACACACACACACACACACACACACACACACAC 1093
QY 4410 AGACAAAGCTGGCAGATTTTCAGAAAAATACAAATTTGCGAATTTGCAGACCTGTCTAC 4469
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1094 CAACAAACACACACACACACACACACACACACACACACACACACACACAC 1153
QY 4470 CGATTATTAACACACACAAATCTCC 4492
QY | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1154 CAAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 1176
XX
XX RESULT 55
XX ABT04778
XX ID ABT04778 standard; DNA; 5163 BP.
XX
XX AC ABT04778;
XX
XX DT 27-SEP-2002 (first entry)
XX
XX DE C parvum GP900 gene fragment SEQ ID NO: 4.
XX
XX KW Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
XX gene; ds.
XX
XX OS Cryptosporidium parvum.
XX
XX PN WO200194631-A1.
XX
XX PD 13-DEC-2001.
XX
XX PF 14-MAY-2001; 2001WO-US15624.
XX
XX PR 06-JUN-2000; 2000US-0588995.
XX
XX PA (REGC ) UNIV CALIFORNIA.
XX
XX PI Petersen C, Barnes DA, Nelson RG, Gut J;
XX
XX WPI: 2002-566447/60.
XX
XX DR Detecting Cryptosporidium in biological and environmental samples and
XX PT diagnosis of cryptosporidiosis involves, contacting the sample with
XX PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
XX
XX PS Disclosure: Page 103-104; 157pp; English.
XX
XX The present invention relates to a method of detecting Cryptosporidium in
XX CC biological and environmental samples, and of diagnosing
XX CC cryptosporidiosis. This involves obtaining a sample and contacting it
XX CC with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
XX CC RNA, or its variant, mutant or fragment. The method is also useful for
XX CC detecting and identifying individual cryptosporidium isolates based on
XX CC the genetic characteristics, and for diagnosis of prior or concurrent
```


CC Cryosporidium infection. The present sequence is a C. parvum coding
CC sequence used in the exemplification of the invention.
XX
SQ Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other;

SQ Sequence 5163 BP; 1873 A; 1138 C; 875 G; 1277 T; 0 other;

Query Match	1.2%	Score 60.6	DB 24	Length 5163
Best Local Similarity	44.2%	Pred. No. 0.00053		
Matches 249; Conservative	0;	Mismatches 314;	Indels 0;	Gaps 0;

QY	3930	CAATGTCACGTAGACGACTTCTGAAACGCGCTTTAATAGAGAAAATTTCAAA	3989
Db	614	CAACAACAACAACAGACACACAAACAACACTRACAACTACTACACTACTACGA	673
QY	3990	CGTCTTGAGAGCGACTGATTCCCTATATAAGACCTGACCTCTCTTAACCTACGTTGTCAA	4049
Db	674	CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	733
QY	4050	GGACCCCTACAACTTCCTCAATTAAGAAGAGGATTAATTTCTGTCTAGTTACAATA	4109
Db	734	CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	793
QY	4110	TATCAAGACTCCATGAGACCGATFATCAATTTGCTAATGATGTGTGGGTATTACAA	4169
Db	794	CAACCAACTACTACACCAACAACAACCAACAACAACAACAACAACAACAACAACA	853
QY	4170	GATCTGAGCGAAAAATACAGCTTACCTTGACTCTATTAAAAAGTATATACAGATA	4229
Db	854	CCAAACTACAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACA	913
QY	4230	GCAAGCGGAGATGAAAAATATGTGCTTCTGTAATACATGAAACCGTGTACAGAC	4289
Db	914	AGAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	973
QY	4290	AGTGAACGACAAATGACCTCTTGCTAATTACCTGGAGGCCAAGTCTTCACTATAC	4349
Db	974	CCAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	1033
QY	4350	TTACGAAAGCAAGTGGAAATGTAATTAATTAAGGAGTGAACCTCAACAAACAATCA	4409
Db	1034	CTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	1093
QY	4410	AGACAAGCTGGCAGATTTCAAGAAAAATTAACAATTTGCTGGAAATGGACAGCTGTAC	4469
Db	1094	CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	1153
QY	4470	CGATTATACCAACAACAATCTCC	4492
Db	1154	CAACTACTACTACTACTACTACACC	1176
RESULT 56			
AAA61848			
ID	AAA61848	standard; DNA; 5318 BP.	
XX	AAA61848:		
AC			
XX			
DT	28-OCT-2000	(first entry)	
XX			
DE	DNA encoding a portion of Cryptosporidium parvum NINC isolate GP900.		
XX			
KM	GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;		
KM	competitive inhibition; attachment; invasion; ligand binding; sporozo-		
KM	merozoite; diarrhoea; protozoan; ds.		
XX			
OS	Cryptosporidium parvum.		
XX			
XX			
FT	Key	Location/Qualifiers	
FT	PH	1669..7182	
FT	CDS	/tag= a	
FT		/partial	
FT		/product= "Cryptosporidium parvum NINC isolate GP900"	
FT		/note= "No start codon given in the specification"	
FT	3'UTR	5167..5318	

```
ET /★tag= b
```

PN US6071518-A.

PD 06-JUN-2000

PF 12-SEP-1997; 97US-0928361.

PR 13-SEP-1996; 96US-0026062.

PR 29-MAY-1992; 92US-0891301.

PR 14-AUG-1996; 96US-0700651.

PA (REGC) UNIV CALIFORNIA.

PI Petersen C;

DR WPI; 2000-422065/36.

[illegible]

PT parvum, useful for c

PT Inhibition of the

PS Claim 16; column 41-48; 59pp; English

CC The invention relates to the GP900 gly

administration of GP900 or fragments thereof to a host to elicit anti-GP900 antibody production, and to a method of cryptosporidiosis treatment or prophylaxis comprising administration of anti-GP900 antibodies to an individual. Cryptosporidium parvum GP900 and GP900 fragments are able to competitively inhibit sporozoite or merozoite attachment or invasion, and are also useful for the generation of anti-GP900 antibodies. The antibodies also inhibit sporozoite or merozoite attachment/invasion and additionally inhibit the binding of GP900 ligands to GP900. GP900 proteins, fragments and antibodies may therefore be used to treat or prevent cryptosporidiosis. Infection with Cryptosporidium is a common cause of diarrhoea in humans and causes life-threatening diarrhoea in immunocompromised persons. Cryptosporidiosis can be contracted from contaminated municipal water supplies (e.g., public swimming pools). It is also a cause of disease in animals, resulting in financial losses in agriculture. GP900 fragments, fusion proteins and antibodies may also be used for the diagnosis of Cryptosporidium parvum infections, and for the detection of the parasite in the environment. The present sequence represents genomic DNA encoding a portion of the GP900 protein of the NMC isolate of Cryptosporidium parvum.

SQ Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 other;

Query Match	1.28;	Score 60.6;	DB 21,	Length 538;
Best Local Similarity	44.28;	Pred. No. 0.00054;		
Matches 249; Conservative	0;	Mismatches 314;	Indels 0;	Gaps 0;

[illegible]

OY 41170 GACCGTAGCGGCAAAATACAGCTGCACCTTGACTCTATTAAAAAGATATTCACACGTAA 4229
 Db 854 CCACAACTACAAACCAACAACTACACCAACAACAACCAACCAACCAACCAACCACTACCA 913
 OY 4230 GCNAGCGCGAATGTAAAAATATCTGCCCTTCCTGAAATTAACATCGAAACCCCTGTACAAGAC 4289
 Db 914 AGAAACCAACCAACCAACCACTACACCAACAACAACAACAACAACAACAACAACAACAACA 973
 OY 4290 AGTGAACGACAAATATGACCTCTTGCTAATTACCTGGAGGCCAAGGTCTCTCACTATAC 4349
 Db 974 CCACACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1033
 OY 4350 TTACGAGGAGACCAATGTGGAAGTTAAATTCAGAGGAGCTGACCTCAAAACAATCCA 4409
 Db 1034 CTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1093
 OY 4410 AGACAAAGCTGGAGATTTTCAGAAAAATACAAATTTGTCGGAATTCAGACAGCTGTCTAC 4469
 Db 1094 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1153
 OY 4470 CGATTATATACCAACAACAATCTCC 4492
 Db 1154 CAAGTACTACTACTACTACTACACC 1176
 RESULT 57
 ABT04777
 ID ABT04777 standard; DNA; 5318 BP.
 XX
 AC ABT04777;
 XX
 DT 27-SEP-2002 (first entry)
 XX
 DE C parvum GP900 gene fragment SEQ ID NO: 3.
 XX
 DE C parvum GP900 gene fragment SEQ ID NO: 3.
 XX
 KM Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
 RW gene; ds.
 XX
 OS Cryptosporidium parvum.
 OS
 PN WO200194631-A1.
 PN
 PD 13-DEC-2001.
 PD
 PF 14-MAY-2001; 2001WO-US15624.
 PF
 PR 06-JUN-2000; 2000US-0588995.
 PR
 PA (RESC) UNIV CALIFORNIA.
 PA
 PI Petersen C, Barnes DA, Nelson RG, Gut J;
 PI
 PT WPI: 2002-566447/60.
 PT
 XX
 DR
 XX
 PS Disclosure: Page 101-103; 157pp; English.
 PS
 CC The present invention relates to a method of detecting Cryptosporidium in
 CC biological and environmental samples, and of diagnosing
 CC cryptosporidiosis. This involves obtaining a sample and contacting it
 CC with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
 CC RNA, or its variant, mutant or fragment. The method is also useful for
 CC detecting and identifying individual Cryptosporidium isolates based on
 CC the genetic characteristics, and for diagnosis of prior or concurrent
 CC Cryptosporidium infection. The present sequence is a C. parvum coding
 CC sequence used in the exemplification of the invention.
 CC
 SO Sequence 5318 BP; 1939 A; 1158 C; 890 G; 1331 T; 0 other;

[illegible]


```

PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation
XX
XX Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention.
XX
XX Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 1 other;
SQ
Query Match 1.2%; Score 60.2; DB 24; Length 6668;
Best Local Similarity 42.2%; Pred. No. 0.00076;
Matches 341; Conservative 0; Mismatches 468; Indels 0; Gaps 0;
QY 672 CTTGCGATATGCGACGCTCTGACACATCAGGACCAATGTGGGAAAGATGAGATTA 731
D 3372 CTACAAATTAACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 3313
QY 732 TATTAATAAGATAAGAACCATCGAGACATTAACAGCTGATCGAAGATCCAAAA 791
D 3312 AAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAACAAAAAACAAC 3253
QY 792 GACCATAGACAAATAATAGATGCAACCAAGAGAGAAAGAAAGATTGTACAGGC 851
D 3252 CAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 3193
QY 852 CCACTAGACCTGTCATCTATACAAACAGCTTGAGAGAGCCATCACTACAGCT 911
D 3192 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 3133
QY 912 ACTGGAGAGCGCATAGACACCTCAGAGAAATGAATATCAAGAGCTGCTGACAA 971
D 3132 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 3073
QY 972 GATTATGAAATTTAAGATCTCTCGCCGCGCAAGCTGTGGAACACCCCTAACAGCTGCT 1031
D 3072 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 3013
QY 1032 GGACAGAGAAAGATAGAGAGAGAGAGAGAGATCAAGAGATGCGCAAAACCAT 1091
D 3012 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 2953
QY 1092 TAACTTCACTAGATCTCTCTTACTGATCCCTTAGCTGGAGTACTCTTGAGAGA 1151
D 2952 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 2893
QY 1152 GAGAGATAGATATAGATCTCTCGCCGCGCAAGCTGAGAGCAAGAGATCAACGAACTTAA 1211
D 2892 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 2833
QY 1212 TGAATATCCCAATGCTGTGAGCTACCTCTCTTATTAACGATATCAACAGCTCTCA 1271
D 2832 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 2773
QY 1272 CGACCTCAATAGCTTGGTGTGATTAACCCCTTGATATATACGAAGAACCTCTTAA 1331
D 2772 CAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 2713

```

```

QY 1332 GAATATCTACACAGCAATGAGAGAGAAAGATTATCAACGAATCAAGAGAAATCA 1391
D 2712 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2653
QY 1392 AATTGAGAGAGAGAAATTTGAGAGTGCACAGAAAAATTACGAGAACCCGCAAAAGTCT 1451
D 2652 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 2593
QY 1452 AAACGATCTCACTAAGAGTATGAAAAAC 1480
D 2592 AAACGAAAAAAGAAAAAAGAAAAAAGAAAAAATC 2564

RESULT 59
AAS00654
ID AAS00654 standard; DNA; 489 BP.
XX
XX AAS00654;
AC
XX 07-SEP-2001 (first entry)
PT
XX
XX Plasmodium vivax merozoite surface protein DNA.
DE
XX
XX Merozoite surface protein; malaria; blood; serum; diagnosis; vaccine; ds;
KW antigen-antibody composite; Enzyme Linked Immunosorbent Assay; ELISA;
KW MSP.
XX
XX Plasmodium vivax.
OS
XX
XX Key Location/Qualifiers
FH 1..489
FT CDS /*tag= a
FT /partial
FT /product= "P. vivax merozoite surface protein"
FT /note= "No start or stop codon"
XX
XX W0200136587-A2.
XX
XX 25-MAY-2001.
PD
XX
XX 15-NOV-2000; 2000WO-KR01302.
PF
XX
XX 15-NOV-1999; 99KR-0050616.
PR 25-APR-2000; 2000KR-0022041.
PR 20-MAY-2000; 2000KR-0027305.
XX
XX (HUMA-) HUMANBIO CO LTD.
XX
XX Park H;
XX
XX WPI: 2001-343809/36.
XX P-PSDB; AAU00668.
XX
XX New gene encoding merozoite surface protein of Plasmodium vivax, useful
XX for producing protein for diagnosis of malaria and for vaccination -
XX
XX Claim 1; Page 21-22; 24pp; English.
XX
XX The sequence represents a DNA encoding a Plasmodium vivax merozoite
XX surface protein (MSP). The C-terminal region of the merozoite surface
XX protein has a strong antigenicity in malarial diseases. For diagnosis of
XX malaria, recombinant proteins with enhanced antigenicity, obtained by
XX addition of fusion proteins to surface protein C-terminal regions, can be
XX reacted with serum or blood of a Plasmodium infected patient.
XX Antigen-antibody composites will be formed, and these are detected by
XX Enzyme Linked Immunosorbent Assay (ELISA). The recombinant antigens
XX provide a quick and reliable diagnosis of malaria, with good sensitivity
XX and selectivity.
XX
XX Sequence 489 BP; 180 A; 134 C; 95 G; 80 T; 0 other;
SQ
Query Match 1.2%; Score 59.6; DB 22; Length 489;
Best Local Similarity 51.9%; Pred. No. 0.00028;

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2003, 00:07:15 ; Search time 125 Seconds
(without alignments)
12119.866 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940

Sequence: 1 cgcacgcgctatgaaatcatc.....ttcatctaatagatcgtg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 180 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1265.2	25.6	5181	1	US-08-257-073-10 Sequence 10, Appl
2	731.6	6.7	1219	4	US-08-195-705-1 Sequence 1, Appl
3	74.2	1.5	7218	1	US-08-232-463-14 Sequence 14, Appl
4	60.6	1.2	5163	3	US-08-700-651-1 Sequence 1, Appl
5	60.6	1.2	5163	3	US-08-928-361B-4 Sequence 4, Appl
6	60.6	1.2	5318	3	US-08-700-651-2 Sequence 2, Appl
7	60.6	1.2	5318	3	US-08-928-361B-3 Sequence 3, Appl
8	58.8	1.2	5511	3	US-08-928-361B-2 Sequence 2, Appl
9	58.8	1.2	7334	3	US-08-928-361B-1 Sequence 1, Appl
10	50	1.0	5661	4	US-08-938-105-2 Sequence 2, Appl
11	49.6	1.0	1430	1	US-08-276-452A-25 Sequence 25, Appl
12	49.6	1.0	1430	1	US-08-798-744-25 Sequence 25, Appl
13	48.2	1.0	1690	1	US-08-276-452A-24 Sequence 24, Appl
14	48.2	1.0	1690	2	US-08-798-744-24 Sequence 24, Appl
15	46.4	0.9	198	5	PCT-US95-10668-3 Sequence 3, Appl
16	46.4	0.9	198	5	PCT-US95-10668-4 Sequence 4, Appl
17	46.4	0.9	954	4	US-08-098-327E-37 Sequence 37, Appl
18	46.4	0.9	954	4	US-08-462-625-37 Sequence 37, Appl
19	46.4	0.9	988	4	US-08-098-327E-34 Sequence 34, Appl
20	46.4	0.9	988	4	US-08-462-625-34 Sequence 34, Appl
21	45.8	0.9	340	5	US-08-182-175A-104 Sequence 104, App
22	45.8	0.9	340	5	PCT-US92-06412-104 Patent No. 5171843
23	45.6	0.9	697	6	5171843-10 Patent No. 5171843
24	45.6	0.9	1137	6	5171843-8 Patent No. 5171843
25	45.4	0.9	198	5	PCT-US95-10668-1 Sequence 1, Appl
26	45.4	0.9	198	5	PCT-US95-10668-2 Sequence 2, Appl
27	45	0.9	2949	4	US-09-412-554A-3 Sequence 3, Appl

28	44.8	0.9	950	4	US-08-098-327E-32 Sequence 32, Appl
29	44.8	0.9	950	4	US-08-462-625-32 Sequence 32, Appl
30	44.8	0.9	1482	4	US-08-098-327E-41 Sequence 41, Appl
31	44.8	0.9	1482	4	US-08-098-327E-45 Sequence 45, Appl
32	44.8	0.9	1482	4	US-08-462-625-41 Sequence 41, Appl
33	44.8	0.9	1482	4	US-08-462-625-45 Sequence 45, Appl
34	44.8	0.9	2223	4	US-08-257-073-4 Sequence 4, Appl
35	44.4	0.9	4766	5	PCT-US93-07261-10 Sequence 10, Appl
36	44.2	0.9	2277	1	US-08-676-967-2 Sequence 2, Appl
37	44.2	0.9	2277	1	US-08-676-974-2 Sequence 2, Appl
38	44.2	0.9	2277	2	US-09-098-487-2 Sequence 2, Appl
39	43.6	0.9	1956	4	US-08-559-896B-1 Sequence 1, Appl
40	43	0.9	1241	1	US-08-471-033-39 Sequence 39, Appl
41	43	0.9	1241	1	US-08-471-033-42 Sequence 42, Appl
42	43	0.9	1241	2	US-08-471-044-39 Sequence 39, Appl
43	43	0.9	1241	2	US-08-471-044-42 Sequence 42, Appl
44	43	0.9	1241	2	US-08-463-483A-39 Sequence 39, Appl
45	43	0.9	1241	2	US-08-463-483A-42 Sequence 42, Appl
46	43	0.9	1241	2	US-08-471-046A-39 Sequence 39, Appl
47	43	0.9	1241	2	US-08-471-046A-42 Sequence 42, Appl
48	43	0.9	1241	2	US-08-470-566B-39 Sequence 39, Appl
49	43	0.9	1241	2	US-08-470-566B-42 Sequence 42, Appl
50	43	0.9	1241	2	US-08-469-334-39 Sequence 39, Appl
51	43	0.9	1241	2	US-08-469-334-42 Sequence 42, Appl
52	43	0.9	1241	3	US-09-300-529-39 Sequence 39, Appl
53	43	0.9	1241	3	US-09-300-529-42 Sequence 42, Appl
54	43	0.9	1358	1	US-08-471-033-45 Sequence 45, Appl
55	43	0.9	1358	2	US-08-471-044-45 Sequence 45, Appl
56	43	0.9	1358	2	US-08-463-483A-45 Sequence 45, Appl
57	43	0.9	1358	2	US-08-471-046A-45 Sequence 45, Appl
58	43	0.9	1358	2	US-08-470-566B-45 Sequence 45, Appl
59	43	0.9	1358	2	US-08-469-334-45 Sequence 45, Appl
60	43	0.9	1358	3	US-09-300-529-45 Sequence 45, Appl
61	43	0.9	1389	1	US-08-471-033-27 Sequence 27, Appl
62	43	0.9	1389	2	US-08-471-044-27 Sequence 27, Appl
63	43	0.9	1389	2	US-08-463-483A-27 Sequence 27, Appl
64	43	0.9	1389	2	US-08-471-046A-27 Sequence 27, Appl
65	43	0.9	1389	2	US-08-470-566B-27 Sequence 27, Appl
66	43	0.9	1389	2	US-08-469-334-27 Sequence 27, Appl
67	43	0.9	1389	3	US-09-300-529-27 Sequence 27, Appl
68	43	0.9	4031	1	US-08-471-033-49 Sequence 49, Appl
69	43	0.9	4031	2	US-08-471-044-49 Sequence 49, Appl
70	43	0.9	4031	2	US-08-463-483A-49 Sequence 49, Appl
71	43	0.9	4031	2	US-08-471-046A-49 Sequence 49, Appl
72	43	0.9	4031	2	US-08-470-566B-49 Sequence 49, Appl
73	43	0.9	4031	2	US-08-469-334-49 Sequence 49, Appl
74	43	0.9	4031	3	US-09-300-529-49 Sequence 49, Appl
75	42.6	0.9	1493	4	US-08-098-327E-38 Sequence 38, Appl
76	42.6	0.9	1493	4	US-08-462-625-38 Sequence 38, Appl
77	42.6	0.9	1995	1	US-08-425-069-3 Sequence 3, Appl
78	42.6	0.9	1995	2	US-08-317-844B-3 Sequence 3, Appl
79	42.4	0.9	2447	2	US-09-014-969-14 Sequence 14, Appl
80	41.8	0.8	774	4	US-09-461-657-187 Sequence 187, App
81	41.8	0.8	819	4	US-09-461-657-185 Sequence 185, App
82	41.6	0.8	1669	4	US-09-461-657-184 Sequence 184, App
83	41.6	0.8	1236	2	US-08-741-134-5 Sequence 5, Appl
84	41.6	0.8	3024	4	US-08-923-992A-3 Sequence 3, Appl
85	41.6	0.8	5312	4	US-09-307-143-1 Sequence 1, Appl
86	41.4	0.8	1399	1	US-08-471-033-24 Sequence 24, Appl
87	41.4	0.8	1399	2	US-08-471-044-24 Sequence 24, Appl
88	41.4	0.8	1399	2	US-08-463-483A-24 Sequence 24, Appl
89	41.4	0.8	1399	2	US-08-471-046A-24 Sequence 24, Appl
90	41.4	0.8	1399	2	US-08-470-566B-24 Sequence 24, Appl
91	41.4	0.8	1399	2	US-08-469-334-24 Sequence 24, Appl
92	41.4	0.8	1399	3	US-09-300-529-24 Sequence 24, Appl
93	41.2	0.8	4411529	4	US-09-103-840B-1 Sequence 1, Appl
94	41	0.8	1360	6	US-08-961-083-37 Sequence 37, Appl
95	41	0.8	1866	6	5210183-1 Patent No. 5210183
96	41	0.8	12665	4	US-08-961-527-134 Patent No. 5180810
97	40.8	0.8	1131	6	5180810-3 Patent No. 5180810
98	40.8	0.8	1784	6	5180810-2 Patent No. 5180810
99	40.8	0.8	2058	2	US-08-749-391-1 Sequence 1, Appl
100	40.8	0.8	2058	3	US-09-390-200-1 Sequence 1, Appl

101	40.8	0.8	3407	1	US-08-255-155A-7
102	40.6	0.8	233	1	US-08-182-175A-86
103	40.6	0.8	233	3	PCT-US92-06412-86
C 104	40.4	0.8	188	1	US-08-115-457-21
C 105	40.4	0.8	188	1	US-08-465-670-21
C 106	40.4	0.8	188	2	US-08-291-011-1
C 107	40.4	0.8	188	4	US-09-266-065-1
C 108	40.4	0.8	405	2	US-08-259-074A-1
C 109	40.4	0.8	405	4	US-09-399-773-1
110	40.2	0.8	9636	1	US-08-332-170B-1
111	40.2	0.8	9636	2	US-08-954-441-1
112	40	0.8	480	2	US-08-726-306A-185
C 113	40	0.8	3095	6	5231168-1
C 114	39.8	0.8	14442	3	US-08-781-891-208
115	39.6	0.8	1266	4	US-09-470-175-3
116	39.6	0.8	1796	4	US-09-470-175-1
C 117	39.6	0.8	4403765	4	US-09-103-840A-2
118	39.4	0.8	392	4	US-09-370-838-18
119	39.4	0.8	2032	4	US-09-241-581B-5
120	39.4	0.8	2032	4	US-08-265-428-5
121	39.4	0.8	2032	5	PCT-US95-07721-5
122	39.2	0.8	720	4	US-09-172-063-17
123	39.2	0.8	720	4	US-09-316-919-18
124	39.2	0.8	768	3	US-09-172-063-33
125	39.2	0.8	3489	2	US-08-728-333A-1
C 126	39.2	0.8	3489	4	US-09-296-568-1
C 127	39.2	0.8	33207	2	US-08-770-379-20
C 128	39.2	0.8	33207	4	US-08-757-669A-20
C 129	39.2	0.8	33207	4	US-09-230-371A-20
C 130	39	0.8	535	4	US-09-056-556-171
131	39	0.8	535	4	US-09-072-556-166
132	39	0.8	696	4	US-09-461-657-193
133	39	0.8	696	4	US-09-461-657-191
134	39	0.8	717	4	US-09-461-657-189
135	39	0.8	828	4	US-09-322-478-17
136	39	0.8	15331	3	US-09-128-155-16
137	38.8	0.8	1695	4	US-09-308-022-5
138	38.8	0.8	3337	1	US-08-072-610-1
139	38.8	0.8	3337	2	US-08-719-822B-1
140	38.8	0.8	3337	4	US-09-092-458-1
141	38.6	0.8	440	4	US-09-397-787-285
142	38.6	0.8	1621	2	US-08-933-750C-96
143	38.6	0.8	1621	3	US-09-234-613-96
144	38.6	0.8	2089	1	US-08-441-139-13
145	38.6	0.8	19124	2	US-08-487-826B-13
146	38.4	0.8	240	1	US-08-628-417-6
147	38.4	0.8	2065	4	US-09-319-989-5
148	38.4	0.8	2394	4	US-09-319-989-9
149	38.4	0.8	3081	4	US-09-319-989-9
150	38.4	0.8	3360	4	US-09-319-989-8
C 151	38.4	0.8	3384	4	US-08-932-992A-5
C 152	38.4	0.8	4200	1	US-08-242-932-1
C 153	38.4	0.8	4200	1	US-08-714-481-1
C 154	38.4	0.8	4200	4	US-08-923-992A-1
C 155	38.4	0.8	4200	5	PCT-US95-06111-1
156	38.2	0.8	1596	4	US-09-134-001C-2083
157	38.2	0.8	2462	4	US-08-847-065-24
158	38.2	0.8	3023	4	US-09-306-022-4
159	38.2	0.8	3394	3	US-08-688-376-1
160	38	0.8	289	4	US-09-007-005-17
161	38	0.8	289	4	US-09-244-706-17
162	38	0.8	321	4	US-09-056-556-159
163	38	0.8	321	4	US-09-072-556-154
C 164	38	0.8	356	2	US-08-520-678A-22
C 165	38	0.8	356	2	US-08-897-126-22
166	38	0.8	765	4	US-08-914-479A-3
167	38	0.8	855	4	US-08-914-479A-5
168	38	0.8	918	3	US-08-937-271-9
169	38	0.8	1051	4	US-09-245-041-10
170	38	0.8	1305	4	US-09-329-234A-6
171	38	0.8	1515	4	US-09-080-625-6
172	38	0.8	1515	4	US-09-695-782-6
173	38	0.8	3153	4	US-09-080-625-3

[illegible][illegible]

D 61 ACACATGAAAGTTATCAAGAACTGTCAAAAACTAGAGCTTTAGACAGTGCAGTATTG 120
O 130 ACCGATACAGCCTGTCCAGAAAGGAGATGCTCAATGAA----- 174
D 121 ACAGGTTATGGTTTATTTCATTAAGGAAAAAATGATCTTAAAGAAAGAAATTACTACA 180
O 175 -----GGAGAGTGGCAGCGCGCTTACAAACGACACCGCT 213
D 181 AAAGTGCAAGTCTCAAAAGTGTACAAAGTGTACAAAGTGTACAAAGTGTACAAAGTGT 240
O 214 TCTAAAGGCTGTGGCTAGCGGTGCTCCGGTGTGTGGCTGTGGCTGTGGCTGTGGCTGT 273
D 241 ACAAGGAGTACAAAGTGTGTCAAGTGTGTCAAGTGTGTCAAGTGTGTCAAGTGTGTCA 300
O 274 GCCTCCGGGGAAGCGGTGCATCAGTGTGTCAAGTGTGTCAAGTGTGTCAAGTGTGTCA 333
D 301 GGTACAAAGTGTGTCAAGTGTGTCAAGTGTGTCAAGTGTGTCAAGTGTGTCAAGTGT 360
O 334 CGAAGAACCAATCATCTGACA----- 355
D 361 AGTCATCATCTGTTCAAACACTTACCTGTCATAATCTCATCTGTGTGCAAGCCCT 420
O 356 -----ACTTAGCGATTCCGACCCAAAGTCTTACGCCGACCTCAAGCAGAGTGA 408
D 421 CCAGCTGATGCAAGCGATTGATGCTTAATCTTACGCTGATTTAAACACAGAGTACGA 480
O 409 AACTATCTCTCACTATCAAGAGTGAAGTACCCACAGTGTGTGACCTCACTATCAT 468
D 481 AATTACTTTTCACTTTTAAAGAACTCAATATCCGAACTCTTGTGATTTAAACAATCAT 540
O 469 ATGTGACACCTGTGTATACATTCATGCTCAAAATATCTGATGACGCTTACGAAGAG 528
D 541 ATGTTAACTTGTGTGATTAATATCATGTTCAAAATATTATATGATGATATGAGAA 600
O 529 ATCAATGAACCTCTGTACAGTGAATTTCTACTGACTGTCTTAAGGCCAAACTGAT 588
D 601 AATTAATGAATTTATTAATTAATTAACCTTTATTTGATTTATTAAGACCAAAATTAAT 660
O 589 GAGCTTGGCCAAATGACTATGTCAAAATTCATTCATTTGAAGTTCAGAGCCAAAGAG 648
D 661 GATGTATGTGTATATGATTTATGTCAAAATCTTCAATCTTAAATTTGTGCAAAATGA 720
O 649 TTGACGATTTAGAGAGTTGGTCTTCCGATATCCGAACTCTGACAAACATCAAGAG 708
D 721 TTAGACGTAATTAATAAACTGTGTGCGATATGAGAAACCAATTAATTAAGAT 780
O 709 AATGTGGGAAAGTGAAGATTTATTAATAAAGATTAAGAACCATCGAAGACATTAAC 768
D 781 AATGTAGGAAAAATGAAAGATTAATTAATAAATAAATAAATAAATAAATAAATAAATA 840
O 769 GAGCTGATGAGAGAAATCCAAAAAGACATGACAAAAATTAAGATTCAGCAAGAGAGAA 828
D 841 GAATTAATTAAGAGAGATTAAGAAACATTTGATCAAAATTAAGATTAAGATTAAGAGAA 900
O 829 GAAAGAGAGAGTTGTACAGGCGCCAGTACGACCTGTCCATTAATTAACAGAGTTGAA 888
D 901 GAAAAAATAAATAATTAACCAAGCTCAATGATCTTTATTTACAAATAAACAATTAGAA 960
O 889 GAAGCCCATTAACCTGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 948
D 961 GAGAGACATTAATTAATAAGCGTTTGAAGAAACGATTTGAACTTTAAAAAATAGAA 1020
O 949 AATATCAAAAGACTGCTGACAAAGATTAATGAATTAAGAACTTCCCGCAGCAACTCT 1008
D 1021 AACATTAAGAGATTTACTGATATGATTAATTAATTAATTAATTAATTAATTAATTA 1080
O 1009 GGAAGACCCCTTAACAGCTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1068
D 1081 GGAAGATACCAAAATACTCTCTGTATAGAACAAAAAATGAGAGAGAGAGAGAGAGAA 1140
O 1069 ATCAAGAGAGATGCGCAAAACATTAAGTTCAACATGATGATCTCTTACTGATCCCTT 1128
D 1141 ATTAAGAGAGATGCGCAAAACATTAATTAATTAATTAATTAATTAATTAATTAATTA 1200

O 1129 GAGCTGAGTACTACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1187
D 1201 GAATTAAGATTAATTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
O 1188 -----GACAAGAGATCAACCGAAGCTTAATGAATATCCCAATGTGTGACGTACCT 1239
D 1261 GATCTCAGAGAAATCTGTTCATTAACCAAGATTCCTTATCCAAATGTGTGTATATCTT 1320
O 1240 CTGTCTTATTAACGATATCAACA-----CGCTCTCAACGAGCTCAATAGCTTGGT 1290
D 1321 TTACCACTCAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
O 1291 GACTTGATTAACCCCTGCTGATTAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350
D 1381 GATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1440
O 1351 GAG 1410
D 1441 GAAAG 1500
O 1411 GAGAGTGAACAG 1470
D 1501 AATCAACAG 1560
O 1471 TATGAAG 1530
D 1561 TATGAAG 1620
O 1531 AACTTCAG 1590
D 1621 GTGATGAAG 1680
O 1591 AATACCTTGCATCTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1650
D 1681 AATTAATTTTCACTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
O 1651 AAGTATATGAG 1710
D 1741 TCATATCTTGAAGATTAATTTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
O 1711 AAGAAATCTATGAAG 1770
D 1801 TATACTTTGAAG 1860
O 1771 GATGAAG 1830
D 1861 AGTGAAG 1917
O 1831 ATCTGAG 1890
D 1918 TCCTTAAG 1977
O 1891 ATTGATGAAG 1950
D 1978 ATGAAG 2037
O 1951 CATGTCCGAG 2010
D 2038 CATGTACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2097
O 2011 AAG 2070
D 2098 AAAAAAG 2155
O 2071 GAG 2130
D 2156 AAG 2215
O 2131 GAGATTAACCGAG 2190
D 2216 AAGAT-GGGGGTCACTCCACACACAGATTAATCCCAATGAGAGAGAGAGAGAGAGAG 2274

QY	2191	GATAGCTGCAAGCTCAGACGACAAGGACGAAGACGACGACGCTCCAGTGGCAGTGGCC	2250
Db	2275	GAACACAGAAAACAGAAAGAAACAGTAGGACACACACACAGGTAACATTAACATTACCA	2334
QY	2251	GTTTCAGAGGCTTAAGGCTCAAGTGGCTTACACCACGAGCTCTGTAAATACAAAGCCAG	2310
Db	2335	CCAAAGAGAGTAAAAGTTGTTGAAATTTCAATATGAAACATAGAGTATATGACATTTCCACA	2394
QY	2311	AATGTCAGCAACTGCGACTACCTTGGAAAGCTCTATGAGTTCCGTAATACATCTTACATC	2370
Db	2395	GCCTTGACAAAAACAGTTTATCTTAAGGAATATAGATGAATTTTAACTAAATCATATATA	2454
QY	2371	TGCCACAATATATCTCGCTCTCTCACACACTATGAGAGAGATTTCTTAACAGTAC	2430
Db	2455	TGTCATTAATATATTTTATAGTATCAAACTGATATGGACAAAAATTTTATAGAGTATAT	2514
QY	2431	AAGATTAACCAAGGAGAGGAGTAAACGTCTCTGTGTGATCCATGAGACCTGGCTGCTC	2490
Db	2515	AATTTTACTCCAGAGAGAGAAAATGAAATTAATAATCATGTGATCCATTGATTTATTTATTT	2574
QY	2491	AATATCCAGAACAAACATCCCGTTATGTATTTCTATGTTCGATGGCTTACAAATTTCTTC	2550
Db	2575	AATATTCAAATATAACATACCTGCTATGTATTCATATATATGATATGACAAATGATTTA	2634
QY	2551	TCTCACTGTTTCATGAGATATATGAGAGAGAGATGCTGCACCTGATTAACCTCAA	2610
Db	2635	CAACATCTCTTTTGTGATATATATCAAAAGGAAATGATTTATTTACATTAACCTAAA	2694
QY	2611	GACAAACACAAGATTAGAACCTCTTGAGAGGAAAGCTTAACAAAGCTGC-----	2657
Db	2695	GAGGAAATACATCAATCAAAAAATTTATGAGAGGACAAAAACAAATTACTGGAACATCT	2754
QY	2658	-----CACCTCTGTTAAAGCTCTCTCTTC	2681
Db	2755	ACATCCAGTCCGTGGAAATACACACGTAATATACGTCAATCCGACAATCAGATATTC	2814
QY	2682	CAGTCCATGCAACCACTGCTCTCTCACACCTCAAGCAA-----	2720
Db	2815	CAAAACCAACAAATCAATATGATCTCTTACCATAATCCCAAAATGATGAGTATCATCT	2874
QY	2721	-----GCCCAAGTAGAGCGCTAAGCAGACGACACCTCTCACTGACCAACCTTAATATCA	2775
Db	2875	GGTCCCTGCTGATGTTGAAGAAATCATGATCCCTTAACAGTATTTGCTATTTAGTACGAT	2934
QY	2776	CTGAACCTGTTTGAAGACATCTGTCTCTCGGCAAGATTAAGACAT---CTACCAAGAA	2832
Db	2935	TTGAAAGGTATTTGTATGTCTCTTAATCTTGGAAATTAACATAAGTACATTCATTTA	2994
QY	2833	CTTATTTGGACAGAAATGCTGCCGAACCTCTACGAGAAATATCTGAAAGACGACGACGA	2892
Db	2995	ACCATTTCTACACACAGATGAGAAAAATTTTATGAGAAATATTTTAAAAAATATGATACC	3054
QY	2893	TTCTATTAACGAGACCTCACTACTCTCGGAAATCTAAAGCCGATATACACTCTTT	2952
Db	3055	TATTTTAAATGATGATTAACAACATTTTCGTAATATCTAATTTCAAAAGTATTTACAGTTG	3114
QY	2953	AACGATGAATCTAAACGTAAAGAGCTGGAAGAGACATCAATTAAGCTGAAGAGCACTG	3012
Db	3115	ACC---GAACACAAAAAAATGCAATTAATATGATGAATTAATAAAATTAAGATATCTTTA	3171
QY	3013	CAACTGAGCTTCGACCTGTCACAAACAGTACAAACTGAACCTGAGAGACTCTTGACAAAG	3072
Db	3172	CAGTTATCATTTGATTTTATTAATAATATTAATTTAAAAATTTGAATGATTTATTTAATAG	3231
QY	3073	AAGAGACAGTGGCGCAGTAAAGATGACGATCAACAATTAAGTCTGTCTGAAGAGAG	3122
Db	3232	AAAAAGACTTGCCAGACCAAAATGCAATTAATAAAACCTTACTTTTATTAAGAAACAA	3291
QY	3133	CTTGAAAGCAACCACTCACTGAAACAATCCGAACAGTACTCAGACATTTCTCACTG	3192
Db	3292	TTTAAGATCAAAAATGGAATTCATTAATTAACCCACATATGTTTACAAAAACTTTCTGT	3351
QY	3193	TTCTTCAACAAGAGAGAGAGCCGAGATGCCGAGACAGAGAACACTCTGTGAGAACAC	3252

Db	3352	TTCTTTACAAAAAAGAGACCTGAATTAGCAGAACTGAAAACACATTTAGAAAACACA	3411
Qy	3253	AAGATCTCTTCMAACACTACAAAGGCGCTCGTCAAGATATTATATGCGGAGCTCTCTCT	3312
Db	3412	AAATATATTATGAAACATTTATTAAGAGACTGTTAATATTATATATGGAATCATCTCCA	3471
Qy	3313	CTGAGACTCTCTCCGAGAGAGCAATCCAGACCGGAGATACCTACGCCAGCTCGAGAAC	3372
Db	3472	TTAAAAACCTTTAAGTGAAGTATCAATTCATCAACAGAGATAAATTAATGCAATTTACAAAAA	3531
Qy	3373	TTCAAGGTCCTGTACAGCTGCAGAGCAAGCGAAGGCAACCTCAACCTGAGGAAG	3432
Db	3532	TTTACAGTATTAAGTAAATTAATGGAACACTCAATGATATTTACATTTAGGAAAGAAA	3591
Qy	3433	AAGCTCAGCTACCTCTTAGCGGAGCTGCATCCTGATCGCCGAGCTCAAGGAATCTT	3492
Db	3592	AAATATATCTCTTCTTTCAGTGGATTTACATCTTTAATTTACGATTTAAAGAACTATTA	3651
Qy	3493	AAGAACAGAACTACACCGGCAATAGCCCAAGCAATATATACAGAGCTGATTAACGA	3552
Db	3652	AAAAATAAAAATTAATACAGGTATTTCTCCAGAGTAAAAATTAAGAAAGTTAAACAACT	3711
Qy	3553	CTGGAATCTTACAAAGAGTCTCGCTCAAGGAACAGATGTCG-----	3596
Db	3712	TTAAAACTTTACGAAAAATTTTCTCTCCAGAGCAAGATTAACACAGTTGTACCTCCACT	3771
Qy	3597	-----CACGTGGTCTGMAATCTGGCTCCGACACACAGTGAAGTCTCA	3641
Db	3772	CACACAGATGTACCTCCATCTCCATTATCTGTAAAGGTAAAGTGAAGTTACAGAGTCCACA	3831
Qy	3642	ACCTTAAGAA-----GCTGCATCTACTATATGTCGAGCCGAGTCCATTAATTAAC	3693
Db	3832	AAAGAGAAACCAAAATACCACTTCAGGCTTTATTATTAACGAATTTACCAACAACTAGTA	3891
Qy	3694	ACATCTCAAGACGTGAGAGATGAGGTGCATGCATCATCATTTGTCCTATCTCGGCGAG	3753
Db	3892	CAATTACAAATATTATGAGCAAGAAATGATTTCTTAGTTGTATTACCCATTTTGGAGAA	3951
Qy	3754	AGCGAGGAGAGCTACGATGACCTCGGCGCAGGTGTCACCGGTGAGGCTGCTACTCTTCC	3813
Db	3952	TCCGAAGATTAATGACGAATATTATGATCAAGTAAGTAACGTGAGGAAGC---AATATCTGTC	4008
Qy	3814	GTGATTTGATACATTTCTGTGCCAAAATTCGAGAACGAATACGAGTGTCTATCTGAACCT	3873
Db	4009	ACAATGGAATTAATCTCTCCAGGATTTGAAATGAATGATGTTATATATTTAAACCT	4068
Qy	3874	CTGGAGAGGCTTATAGGCTCTCCAGAAACAGCTGGAATTAATGAGTGAAGCTTCAAT	3933
Db	4065	TTAGCTGGAGTATATAGAAAGCTTAAAAAAAACAATTTGAAAAAACAATTTTTCATTTAAT	4128
Qy	3934	GTCAACGTGAAGACATTTCTGAAACAGCGCTTTAATTAAGAGAAATTTCAAGAACGTC	3993
Db	4129	TTAAATTTTGAAGATATCTTAATTTACAGCTCTTAAGAAACGAATATTTCTTAGATGTA	4188
Qy	3994	TTGGAGAGCGACTGATTTCCCTATTAAGACCTGCAGCTCTCTTAAGTGTTCGAAGAC	4053
Db	4189	TTAGATCTGATTTATATGAATTAATTAACATTAATCTCTCAAAATGAAATTAATTTGAAGAT	4248
Qy	4054	CCATCAAGTCCCTCAATTAAGAGAAAGGGATTAATTTCTGTAGTTACAGCATATATC	4113
Db	4249	TCATTTAAATTTTGAATTTCAACAAAAAACACACTTTTAAAAAGTTACAAATATATA	4308
Qy	4114	AAGGACTCCATGACACCGATATCAATTTCCGTAATGATGTGCTGGGGTATTACAAATC	4173
Db	4309	AAAGATCATGTAAGAAATGATTTAAATTTGGACAGGAAGATTAAGTATTTATGAAAG	4368
Qy	4174	CTGAGCGAAAAATACAAAGTCTGACCTTGACTATTTAAAAAGTATATCAC-----	4224
Db	4369	GTTTATAGCAAAATTAAGGATGATTTAGATCAATTTAAAAAGTTATCAAAAGAAAG	4428
Qy	4225	-----GATAAGCA	4233

Db 4429 GAGAGATTCATCATCAACCAACACACCTCCGTACAGCAAAAACAGACGACAA 4488
QY 4234 GCGGAGATGAAATATCTGCGCTCCGATATACATCGAAGACCTGTACAGACAG 4293
Db 4489 AAGAAGAAAGTAAAGTCTCCATTTTAAACAAACATTGAGACCTTAATACAACTTA 4548
QY 4294 AACGAAATAGCATCTCTGCTAATTCACCTGAGGAGCCAGGCTCTCACTATCTTAC 4353
Db 4549 GTTAAATGAAATGACGATTACTTAATTAACCTTAAGGCAAAATTAACGATTGTAT 4608
QY 4354 GAGAAAGCATGTGGAAGTTAAATCAAGAGCTGAACCTCCATAAACAATCCAGAC 4413
Db 4609 GAAAGATGAGACACATGTTAAATTAATACTTAAGATTGATTTAAACCAATGATGAC 4668
QY 4414 AAGCTGGCGATTTCAAGAAATTAACATTTCTGCGAATTCGACAGCTGTCTACCGAT 4473
Db 4669 AAAATGATCTTTTAAACCATTAACGCTTCGAAAGCAATTAATAATGATTAATGAT 4728
QY 4474 TATTAACCAACATCTCTGACCAAGTTCTGTCACATGCGATGGTGTTCGAAAACCTC 4533
Db 4729 GATAGGAAAAAGATATGCTTGCAAAATTAAGTACGATGATGATGATGATGATGAT 4785
QY 4534 GCGAAACAGCTCTGACATCTGCTGACGCGCAACCTGACAGGAGATGCTGAACATCTCC 4593
Db 4786 CCTAATACATATATCAAAATTAATGGAAGAAATTCAGATATGTTAAACATTTCA 4845
QY 4594 GAGCAACATGGGTGAAGAAACAGTCCCGCAATAGGCGGTTCAGGATCTGAGAC 4653
Db 4846 CACACCAATGGCTTAAACCAATGCTGCAAAATTCGATGTTTCGACATTTAGAT 4905
QY 4654 GAGCGGAGAGTGCAGAGTCTCTGCAACCAACAGAGAGAGATGAGTGCCTGAG 4713
Db 4906 GAAAGAGAGATGTAATGTTTAAATTAATTAACAAAGAGATGATGATGATGATGAT 4965
QY 4714 AACCAACCTTACCTGCAATGAACCAATGCGGGGTGAGCCGAGCTTAATGACACC 4773
Db 4966 AATCCAAATCTTCTGTAACCAAAATATGATGATGATGATGATGATGATGATGATGAT 5025
QY 4774 GAGGAAGACAGGCTCTAAGGAAAGAAATACATGATGATGATGATGATGATGATGATGAT 4833
Db 5026 GAAAGAGATTCAGTACGACGAAAGAAATGATGATGATGATGATGATGATGATGATGAT 5085
QY 4834 TATTCATCTCTGACGAGATTTTTCCTCAGCTTAATTTCTGAGCATCTCTCTCTCTG 4893
Db 5086 TATTCATCTCTGAGATGATTTTTCGAGCTCTCCTCACTTTTGAAGATATCTCTTA 5145
QY 4894 CTGATCTCTGATCTCTGATCTCTGATGATCTCTGATGATCTCTGATGATCTCTGAT 4929
Db 5146 TTAATATCTCATGTTAATATTAATATGATGATGATGATGATGATGATGATGATGAT 5181

RESULT 2
US-08-195-705-1
; Sequence 1, Application US/08195705
; Patent No. 6420523
; GENERAL INFORMATION:
; APPLICANT: Chang, Sandra
; APPLICANT: Hui, George
; APPLICANT: Bair, Philip
; APPLICANT: Gibson, Helen
; TITLE OF INVENTION: BACULOVIRUS PRODUCED PLASMIDIUM
; TITLE OF INVENTION: FAICIPARUM VACCINE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Davis Hoxie Faithfull Hapgood
; STREET: 45 Rockefeller Pl.
; CITY: New York
; STATE: N.Y.
; COUNTRY: USA
; ZIP: 10111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/195,705
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jacobs, Seth H
REGISTRATION NUMBER: 32140
REFERENCE/DOCKET NUMBER: 11880A3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-757-2200
TELEFAX: 212-586-1461
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Plasmidium faiciparum
STRAIN: faiciparum uganda palo alto
IMMEDIATE SOURCE:
CLONE: f195RUP42ACNPV
FEATURE:
NAME/KEY: exon
LOCATION: 13...1194
US-08-195-705-1

Query Match 6.7% Score 331.6; DB 4; Length 1219;
Best Local Similarity 57.9%; Pred. No. 7.3e-78;
Matches 669; Conservative 0; Mismatches 424; Indels 63; Gaps 2;

QY 3765 CTACATGACCTCGGCGGAGGTGTCACCGGTGAGCTGTCACTCTCCGTGATTGATTA 3824
Db 39 CTGGGCTGTCTGTGTCACGACGACACTGTGACCGGGGAGATATCTGCACATGATGA 98
QY 3825 CATTTGTCCAAATGAGAGCAATAGCAAGTGTCTATCTGAAACCTTGGCAGGCGT 3884
Db 99 TATCTCTCAGGATTTGAAATGAAATGATGATGATGATGATGATGATGATGATGATGAT 158
QY 3885 CATATGCTCTCAGAAACACCTGAGATAGCTGATGATGATGATGATGATGATGATGATGAT 3944
Db 159 ATATAGAGCTTAAACCAAAATGAAACCAATTTTACATTTAATTAATTAATTTGAA 218
QY 3945 GACATTTCTGACACGCGCTTAATAGAGAAATTTCAAGACGCTTGGAGAGCA 4004
Db 219 CGATATCTTAATTTACGCTCTTAAGAAACGAAATTTTCTTAGATGATGATGATGATGAT 278
QY 4005 CTGATCTCTTAATAGAGCTGACCTCTTACATGATGATGATGATGATGATGATGATGATGAT 4064
Db 279 TTTAATGCAATTTAAACATATATCTCAAAATGATGATGATGATGATGATGATGATGATGAT 338
QY 4065 CCTCATTAAGAGAGAGGATTAATTTCTGTAGTACATCTATCAAGACTCCAT 4124
Db 339 ATTGAATTCGAAACCAAAACACACTTTTAAAGTTTAAACAAATATTAATAAAGATTCAGT 398
QY 4125 CGACACCGATATCAATTTGCTAATGATGATGATGATGATGATGATGATGATGATGATGAT 4184
Db 399 AGAAATGATATTAATTTTGCACAGAGATTAAGTTATTTGAAAAGTTTATGCGAA 458
QY 4185 ATACAAGTCTGACCTTGAATTAATAAGTATTAACAAGTAAAG----- 4230
Db 459 ATATAGAGTATTAAGATCAATTAATAAAGTATTAACAAGTAAAGAGAGAGATTCCTCC 518
QY 4231 -----CAAGCGAATGA 4244
Db 519 ATCATCACCAACACACCTCCGTACACCAAAACAGAGCAAAAGAGAGAG 578
QY 4245 AAAATATCTGCCCTTCTGATTAACATGCAAAACCTGTACAGAGATGAGACAAAT 4304

DB 579 TAAGTTCCTTCATTTTAAACAACATTTGAGACCTTATACAAATACCTAGTTATAAAT 638
QY 4305 CGACCTCTTGTAATTAACCTGAGGCGCAAGGCTCTACTATACCTAGCAAGAACAA 4364
DB 639 TGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 698
QY 4365 TGAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4424
DB 699 AGCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 758
QY 4425 TTTCAAGAAAATTAACAAATTTGCTGGAATTTGACAGCTGTCTAACCGATTAAACCAA 4484
DB 759 TTTTAAAAACCAATACAGCTTCGAAAGCAATTTAAAAATTTGATTAATGATATACGAAAA 818
QY 4485 CATCTCCCTACCAAGTTTGTGTCACATGCGATGTTTGGAAAACTCCGCAAAACAGT 4544
DB 819 AGATATGCTTGCGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 875
QY 4545 GCTGAGCAATCTGCTCGACGCAACCTGACAGGCACTGTAACATCTCCAGCAACATG 4604
DB 876 AATATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 935
QY 4605 CGTGAAGAAACAGTCCCGCAAGATTAAGCGCTGTTTCAGCATCTGACGACGCGAAGA 4664
DB 936 CGTAAAAAAACAAATGTCAGAAAATTTCTGATGATTTTACAGATTTAGATGAAAGAGA 995
QY 4665 GTCAACAGTCTCTCGTAACAAACAAGAGAGATTAAGTGGCTGAGAGAACCAACCC 4724
DB 996 ATGTAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1055
QY 4725 TACCTGCAATGAACAAATGCGGGGTGACGCCGATGCTAAATGACGACGAGAACAG 4784
DB 1056 TACTTGAACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1115
QY 4785 CGGCTTAACGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4844
DB 1116 AGTAGCAAGGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1175
QY 4845 CGACGAGATTTTTCG 4860
DB 1176 CGATGATTTTTCG 1191

RESULT 3
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ. ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pl-FLS
US-08-232-463-14

Query Match 1.5%; Score 74.2; DB 1; Length 7218;
Best Local Similarity 7.8%; Pred. No. 3,2e-09;
Matches 34; Conservative 234; Mismatches 167; Indels 0; Gaps 0;

QY 1693 AAGAACTAAAGTATTAACAAGATCTCATAGTAAGATGCAAAAAGATGAGAGCTT 1752
DB 1465 AAGTAACTAAAGTATTAACAAGATCTCATAGTAAGATGCAAAAAGATGAGAGCTT 1406
QY 1753 GTTGAGAACATTAAGAAAGTGAAGAACAGTGTGTTGAGAAAGATTACAAAAGCAA 1812
DB 1405 RRR 1346
QY 1813 AATAAACAGATGAGAAGATCTGAGCTCCGATATTGTTAAAGTCAAGTGCAGAG 1872
DB 1345 RRR 1286
QY 1873 GTGCTCTCATGAACAAGATTGATGAACCTAAGAAGCTCACTGATTCGAAAGAGT 1932
DB 1285 RRR 1226
QY 1933 GAGTTAAACATATATATATATGTCGCAATAGTTATTAAGCAGAGAAATTAAGCAACA 1992
DB 1225 RRR 1166
QY 1993 TACTACCTCATGTAACAAGAAAGATGACAACTGAAGTGTTCATGCCAAAGTC 2052
DB 1165 RRR 1106
QY 2053 GAGAGCTGATCAAGAGAGAGAAACATTAATACTGAAGCAGTCAGATACTC 2112
DB 1105 RRR 1046
QY 2113 GAGCTTCACAGAA 2127
DB 1045 CAGCAAGCTCGAA 1031

RESULT 4
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882

GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751

EARLIER FILING DATE: 1995-04-03
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 1
LENGTH: 5163
TYPE: DNA
ORGANISM: Cryptosporidium parvum
US-08-700-651-1

Query Match 1.2%; Score 60.6; DB 3; Length 5163;
Best Local Similarity 44.2%; Pred. No. 1e-05;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

QY 3930 CAATGTCAGTGAAGGAGCATTTCTGACACGCCGCTTTAATAGAGAGAAATTTCAAGAA 3989
DB 615 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 674
QY 3990 CGTCTTGGAGAGCGAGCTGATTTCCCTATTAAGACCTGACCTCCTTAACCTGTTGTCAA 4049
DB 675 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 734
QY 4050 GGACCCATCAAGTTCCTCAATTAAGAGAGAGGTAATTTCTGTCTAGTTACACTA 4109
DB 735 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 794
QY 4110 TATCAAGGACTCCATCGACACCGATATCAATTCGTAATGATGCTGGGATTACAA 4169
DB 795 CAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 854
QY 4170 GATCTGAGCGAATAATACAGTCTGACCTTACTCTATTAAGATATATACAGATAA 4229
DB 855 CCACAACACTACACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 914
QY 4230 GCAAGCGGAGATGAAAAATATCTGCCCTCTGTAATACATGCAAACTTACACAGC 4289
DB 915 AGAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 974
QY 4290 AGTGAAGCAAAATGACCTCTTGTATTCACCTGGAGGCAAGGCTCCTCAATATAC 4349
DB 975 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1034
QY 4350 TTACGAGAAGCAATGTGGAAGTTAAATCAAGGAGCTGAACCTCACTCAAAACAATCA 4409
DB 1035 CTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1094
QY 4410 AGACAAGCTGGCAGATTTCAAGAAATTAACAATTTGTCGGAATTTGACAGACTGTCTAC 4469
DB 1095 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1154
QY 4470 CGATTATAACCAACAACAATCTCC 4492
DB 1155 CAACTACTACTACTACTACTACACC 1177

RESULT 5
US-08-928-361B-4
Sequence 4, Application US/08928361B
Patent No. 6071518

GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: PETERS, VERNY, JONES & BIRSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928.361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4

Query Match 1.2%; Score 60.6; DB 3; Length 5163;
Best Local Similarity 44.2%; Pred. No. 1e-05;
Matches 249; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

QY 3930 CAATGTCAGTGAAGGAGCATTTCTGACACGCCGCTTTAATAGAGAGAAATTTCAAGAA 3989
DB 614 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 673
QY 3990 CGTCTTGGAGAGCGAGCTGATTTCCCTATTAAGACCTGACCTCCTTAACCTGTTGTCAA 4049
DB 674 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 733
QY 4050 GGACCCATCAAGTTCCTCAATTAAGAGAGAGGTAATTTCTGTCTAGTTACACTA 4109
DB 734 CGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 793
QY 4110 TATCAAGGACTCCATCGACACCGATATCAATTCGTAATGATGCTGGGATTACAA 4169
DB 794 CAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 853
QY 4170 GATCTGAGCGAATAATACAGTCTGACCTTACTCTATTAAGATATATACAGATAA 4229
DB 854 CCACAACACTACACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 913
QY 4230 GCAAGCGGAGATGAAAAATATCTGCCCTCTGTAATACATGCAAACTTACACAGC 4289
DB 914 AGAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 973
QY 4290 AGTGAAGCAAAATGACCTCTTGTATTCACCTGGAGGCAAGGCTCCTCAATATAC 4349
DB 974 CCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1033
QY 4350 TTACGAGAAGCAATGTGGAAGTTAAATCAAGGAGCTGAACCTCACTCAAAACAATCA 4409
DB 1034 CTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1093
QY 4410 AGACAAGCTGGCAGATTTCAAGAAATTAACAATTTGTCGGAATTTGACAGACTGTCTAC 4469
DB 1094 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1153
QY 4470 CGATTATAACCAACAACAATCTCC 4492
DB 1154 CAACTACTACTACTACTACTACACC 1176

RESULT 6

QY	4350	TTACGACAAAGAGATGTGGAGTTAAATCAAGGAGCTACCTCAAAACAATCA	4409
Db	1034	CTACCAACAAACAAACACTACTACTACCAACAACAACAACACTACTACTACTA	10933
QY	4410	AGACAACCTGCAGATTTTCAGAAAAATAACAATTCGTGGGAATTTGAGAGCTGTATC	4469
Db	1094	CAA	1153
QY	4470	CGATTATAACCAACAACATTCG	4492
Db	1154	CACTACTACTACTACTACACACC	1176

RESULT 8

```

: US-08-928-361B-2
: Sequence 2, Application US/08928361B
: Patent No. 6071518
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INJECTIONS
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: City: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verny, Hana
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5511 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: OS-08-928-361B-2

```

Query Match	1.2%;	Score 58.8;	DB 3;	Length 5511;
Best Local Similarity	45.2%;	Pred. No. 3.2e-05;		
Matches 216; Conservative	0;	Mismatches 262;	Indels 0;	Gaps 0;

Oy	3336	CACGCTCAGAGCAATTCGTGAACGCCGGCTTTAATAGAGAGAATAATTCCAAAGCGTCT	3995
Dd	1094	CCAGACAACACTACACACACCACACACTACAGAAAACCAACACACACACACACACAA	1153
Oy	3396	GGAAGGAGCATTTATTCCTATTAAGACCTGACATCTCTAACCTACGTTGCAGAGACC	4055
Dd	1154	CACCTTACTACTACTACACACACACACACACAACTACTACTACTACTACTACTACTACTA	1213
Oy	4056	ATACAACTTCCTCAATTAAGAGAGAGGAGTAAATTTCTGCTAGTGTACCAACTATATCAA	4115

D _b	1214	CTACTACTACTACTACTACCAACAACAACAACCACAACCAACCAACTACCCAGAAGAACAA	12737
O _y	4116	GGACTCATCGACACCGATATCAATTTCGTAATGATGTGCTGGGGTATTACAGATCTC	4175
D _b	1274	CACACAAACAACAACAACAACAACAACACTACTCTCAACACGACGAACAACAACACG	13338
O _y	4176	GAGGAAAAAATACAGTGCTGACCTGTGCCTTTATAAAGATATACACGATTAACCAAG	4235
D _b	1334	CACCCACACCCACTACTACACACTACCAAGAACCAACAACACTCTACTACTACTACCA	13939
O _y	4236	CGAATATGAAAAATATCTGCCCTTCTCGATATACATCGAAACCCCTGACAAACAGTAA	4295
D _b	1394	CACACACACTACTACTACTACTACCAACAACAACAACAACACTACTACTACTACCAACA	1453
O _y	4296	CGACAAAATGACCTCTTTCSTAAATTACCTCGAGGCCAAGGTCTCAACTATATTACGA	4355
D _b	1454	CCACACACACACACGCAACCAACAACACTACCAAGAAACCAACAACAACAACACTACTA	1513
O _y	4356	GAAAGCAATGTGGAAATTTAAATCGAAGGCTGAATCTCTCAAAACCAATCTCAAGC	4413
D _b	1514	CTTACTACTACACCAACAACAACAACAACACTTACCACTGCTCCCAACAACAACACTTACT	1571

RESULT 9

```

US-08-928-361B-1
: Sequence 1, Application US/08928361B
: Patent No. 6071518
:
: GENERAL INFORMATION:
:
: APPLICANT: Petersen, Carolyn
:
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
:
: NUMBER OF SEQUENCES: 30
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
:
: STATE: CA
:
: COUNTRY: USA
:
: ZIP: 94306-1840
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/928,361B
:
: FILING DATE: 12-SEP-1997
:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 60/026,062
:
: FILING DATE: 13-SEP-1996
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Verny, Hana
:
: REGISTRATION NUMBER: 30,518
:
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 650-324-1677
:
: TELEFAX: 650-324-1678
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 7334 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: US-08-928-361B-1

```

Query Match	1.28;	Score 58.8;	DB 3;	Length 7334;
Best Local Similarity	45.28;	Pred. No. 3.8e-05;		
Matches 216; Conservative	0;	Mismatches 262;	Indels 0;	Gaps 0;

LENGTH: 1430 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1312
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112
OTHER INFORMATION: /note= "Derived amino acid sequence
OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112
OTHER INFORMATION: /note= "Derived amino acids 14-24,
OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
OTHER INFORMATION: by protein microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 25..31
OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
OTHER INFORMATION: 26 can be T instead of A"
US-08-276-452A-25

Query Match 1.0%; Score 49.6; DB 1; Length 1430;
Best Local Similarity 44.3%; Pred. No. 0.004;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

OY 3085 GGCAGTATTAAGATGATGATCAAGAGTTGACTGCTCAAGAGAGAGCTTGAAGCANA 3144
DB 521 GGCTACTCGGAGATTAACACACACAGAACCAATGGCTACTCGGAGATTACACACAC 580
OY 3145 CTCACCTCAGTACAGCAATCCGAACAGTACGACAACTTCTGTTCTTCAACAG 3204
DB 581 AACACACACATGCTACTCCGAGATTACACACACACAAATGGCTACTCCGAGATT 640
OY 3205 AAGAAGAGCCGAGATGCGCCGAGACAGAACACTCTGAGAACACCAAGATTCTTC 3264
DB 641 TACACACACATGCTACTCCGAGATTACACACACAAATGTTACTCCGAGATTACATG 700
OY 3265 AACACTACAAAGGCTGCTCAAGTATTATTAATGGGAGTCTTCCTGGAAGCTC 3324
DB 701 AACACACACAGGCTCTCCGAGATTACACACACACACACACATATTATTTTC 760
OY 3325 TCCGAGAGAGCATCCAGACGAGATTAATAGCCAGCTCGAGAACTTCAAGTCTG 3384
DB 761 TCCGAGATTACACACACATTAATATGTTTCTCCGAGATTACACACACAT 820
OY 3385 TCTAAGCTGAGAGCAAGCTGAAGCAACCTGAACTGAGAACAGAGCTCAGCTAC 3444
DB 821 AATAACAAATTAATGTTTCTCCGAGATTACACACACATTAATGCTTCTCCGAG 880
OY 3445 CTCCTAGGAGAGCTGATCAGCTGATCCGAGCTCAAGAGTCAATTAAGAAGAAC 3504
DB 881 AACTACACACATTAATGTTTCTCCGAGATTACACACACAAACATTAACATGCT 940
OY 3505 TACACCGGCAATAGCCCAAGCAGATTAATACGAC 3540
DB 941 TTCTCTGAGATTACACACAAACAAACATGCTCC 976

RESULT 12

US-08-798-744-25
Sequence 25, Application US/08798744
Patent No. 5830747

GENERAL INFORMATION:

APPLICANT: Chen, Chao-Guang
APPLICANT: Mau, Shiao-Lim
APPLICANT: Du, He

APPLICANT: Gane, Allison M
APPLICANT: Bacic, Antony
APPLICANT: Clarke, Adrienne E
TITLE OF INVENTION: Plant Arabidogalactan Protein (AGP) Genes
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winer, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: United States of America
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/798,744
FILING DATE: 13-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/276,452
FILING DATE: 18-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Caruthers, Jennie M.
REGISTRATION NUMBER: 34,464
REFERENCE/DOCKET NUMBER: 27-91A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)499-8080
TELEFAX: (303)499-8089
TELEX: 49617824
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1430 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1312
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112
OTHER INFORMATION: /note= "Derived amino acid sequence
OTHER INFORMATION: corresponding to the peptide sequence by protein
OTHER INFORMATION: microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 41..112
OTHER INFORMATION: /note= "Derived amino acids 14-24,
OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
OTHER INFORMATION: by protein microsequencing"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 25..31
OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
OTHER INFORMATION: 26 can be T instead of A"
US-08-798-744-25

Query Match 1.0%; Score 49.6; DB 2; Length 1430;
Best Local Similarity 44.3%; Pred. No. 0.004;
Matches 202; Conservative 0; Mismatches 254; Indels 0; Gaps 0;

OY 3085 GGCAGTATTAAGATGATGATCAAGAGTTGACTGCTCAAGAGAGAGCTTGAAGCANA 3144
DB 521 GGCTACTCGGAGATTAACACACACAGAACCAATGGCTACTCGGAGATTACACACAC 580
OY 3145 CTCACCTCAGTACAGCAATCCGAACAGTACGACAACTTCTGTTCTTCAACAG 3204
DB 581 AACACACACATGCTACTCCGAGATTACACACACAAATGGCTACTCCGAGATT 640

QY 3205 AAGAAGAACCCGAGATCGCCGAGACAGAGAACACTCTGAGAAACACCAGATTCTTCTC 3264
 Db 641 TACAAACAACATGGCTCTCCAAAAAATGCAACATAATGTTACTCCGAAATTTACATG 700
 QY 3265 AAACACTACAAAGGCTTCGTCAGTATTATTAATGCGAGTCTTCTCTGAGACTCTC 3324
 Db 701 AACAAACAACAGCGCTTCTCCGAGATTACACAGCAACAACAACAATATTCTTC 760
 QY 3325 TCCGAGAGAGCATTCAGACCGAGATTAATACGCCCTCGAGACTTCAAGGCTCTG 3384
 Db 761 TCCGAGATTACACACAAATATTAATATATTTCTTCCGAGATTACACACAAAT 820
 QY 3385 TCTAAGCTCGAAGCGAAGCTGAAGACACCTGAACCTGAGAGAAAGAACTCAGCTAC 3444
 Db 821 AATTAACAATAATGTTTCTCCGAGAAATTACAAACAATAAACAACATGCTTCTCCGAG 880
 QY 3445 CTCTCTAGCGGACTGCTACCTCGATGCGGAGCTCAGAGAGTCTTAAGAACACAGAC 3504
 Db 881 AACTTACAAACAATAATGTTTCTCCGAGATTACAAACAACAATAAACAACATGCT 940
 QY 3505 TACACCGGCAATAGCCCAAGCGAGAAATATACAGAC 3540
 Db 941 TTCTCTGAGATTACAAACAACAACAACAATGCC 976

RESULT 13
 US-08-276-452A-24
 : Sequence 24, Application US/08276452A
 : Patent No. 5646029
 : GENERAL INFORMATION:
 : APPLICANT: Chen, Chao-Guang
 : APPLICANT: Mau, Shiao-Lim
 : APPLICANT: Du, He
 : APPLICANT: Gane, Allison M
 : APPLICANT: Clarke, Adrienne E
 : TITLE OF INVENTION: Plant Arabinogalactan Protein (Agp) Genes
 : NUMBER OF SEQUENCES: 91
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Greenlee and Winner, P.C.
 : STREET: 5370 Manhattan Circle, Suite 201
 : CITY: Boulder
 : STATE: Colorado
 : COUNTRY: United States of America
 : ZIP: 80303
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/276,452A
 : FILING DATE: 18-JUL-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Caruthers, Jennie M.
 : REGISTRATION NUMBER: 34,464
 : REFERENCE/DOCKET NUMBER: 27-91A
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (303)499-8080
 : TELEFAX: (303)499-8089
 : TELEX: 49617824
 : INFORMATION FOR SEQ ID NO: 24:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1690 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 60..1442

: FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: 1..38
 : OTHER INFORMATION: /note= "Nucleotide sequence
 : OTHER INFORMATION: obtained by PCR which does not overlap with the
 : OTHER INFORMATION: cDNA clone"
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: 60..128
 : OTHER INFORMATION: /note= "Predicted transmembrane
 : OTHER INFORMATION: segment"
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: 135..179
 : OTHER INFORMATION: /note= "Derived amino acid sequence
 : OTHER INFORMATION: corresponding to the peptide sequence by protein
 : OTHER INFORMATION: microsequencing"
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: 135..179
 : OTHER INFORMATION: /note= "Amino acids 27 to 36, 38,
 : OTHER INFORMATION: and 40 are identical to that in the peptide
 : OTHER INFORMATION: obtained by direct microsequencing"
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: 135..179
 : OTHER INFORMATION: /note= "Amino acid 26 may also be
 : OTHER INFORMATION: determined by direct microsequencing"
 : OTHER INFORMATION: Aln: 37 and 39 can also be undetermined residues"
 : US-08-276-452A-24

Query Match 1.08; Score 48.2; DB 1; Length 1690;
 Best Local Similarity 43.28; Pred. No. 0.01;
 Matches 230; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 859 GACCTGTCATCTATACAAACAGCTGAGAGAGCCCATACCTCAGCTACTGAG 918
 Db 573 GATGAGCTCTTACATATCAATATGTTATGATTCACACACACGACCAACAATGTTAT 632
 QY 919 AAGCGCATAGACCCCTCAAGAAATGAATATCAAGAACTGCTCGACAGATTAT 978
 Db 633 GATTCACAATAATACAAACAAATACAGATGATGCTTCTCCGAGATTACAAACAAT 692
 QY 979 GAATTTAGAAATCTCCGCCAGCCCACTGTGGACACCCCTTAACAGCTGTGGACAG 1038
 Db 693 GGCTACTCGGAGATGCTATTAACAAAAATACAAATGCTACTCAGAAATTACACAAAC 752
 QY 1039 AACAAAGATAGAGGAGCAGCAAGAAAGATCAAGATCGCAAAACCATTAAGTTC 1098
 Db 753 AATTAACAACAATGCTTAGCCCAAGAAATTAACAACAATGCTACTCTCAGACTTACAAAC 812
 QY 1099 AACATAGATTCTCTTTTACTGATCCCTTGAGCTGAGTACTTGAAGAGAAAT 1158
 Db 813 AACAAATATTTTTACTGAGAAATTAACAACAACAACAATATGTTTCTCGAGAAAT 872
 QY 1159 AAGAAATATAGACATCTCCGCCAAGTCCGACAAAGAAATCAACCACTTAATGAATAT 1218
 Db 873 TCCAAACAACAATGCTACTCCAAAAAGATCAACAATATGCTACTCCGAAATTCATG 932
 QY 1219 CCAATGCTGAGAGTACCTCTGTTTATAGATTAACAAGCTCTCAAGGAGCTC 1278
 Db 933 AACAAACAACAATGCTTCTCCGAGAGTTTACAAACAACAACAACAACAACAACAAC 992
 QY 1279 AATAGCTTCGAGTGTGATTAACCCCTTGATTAATCGAAGAAACCTCTAAGAAATATC 1338
 Db 993 AAGCTTTTCTCTGAGATTACAAACAACAATAATATATATGTTTCTCGAAGATTAC 1052
 QY 1339 TACACAGACATGAGAGAAAGAACTTATACGAATTCAGAGAGAAATCAAA 1391
 Db 1053 AACAAACAACAATAACAACAATGCTTCTACGAGAAATTACAAACAACAACAACA 1105

RESULT 14
 US-08-798-744-24

FILED DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
PCT-US95-10668-3

Query Match 0.9%; Score 46.4; DB 5; Length 198;
Best Local Similarity 52.6%; Pred. No. 0.0085;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1357 AAGAGTTTATCAAGCAATCAAGAGATCAAAATTGAGAGAGAAATTTGAGAGT 1416
DB 2 AACACAAAC 61
QY 1417 GACAGAAAGTTACGAGACGCGACGACAAAGTCTAAAGATATCACTAAAGATATGAA 1476
DB 62 AACACAAAC 121
QY 1477 AAGCTGCTCAAGCAGATCTATGATTCCAAATTCACATATACATGACCTGACCACTTC 1536
DB 122 AACACAAAC 181
QY 1537 GAGAAATGATG 1548
DB 182 AAGCTGATCTG 193

RESULT 16
PCT-US95-10668-4
Sequence 4, Application PC/TUS9510668
GENERAL INFORMATION:
APPLICANT: James Eberwine
TITLE OF INVENTION: A Method of Sequencing Proteins by
TITLE OF INVENTION: Epitope Ordering and Protein
TITLE OF INVENTION: Restriction Mapping
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jane Massey Licata, Esq.
STREET: 210 Lake Drive East, Suite 201
CITY: Cherry Hill
STATE: NJ
COUNTRY: USA
ZIP: 08002
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10668
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/294,133
FILING DATE: August 22, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jane Massey Licata
REGISTRATION NUMBER: 32,257
REFERENCE/DOCKET NUMBER: PENN-0137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 779-2400
TELEFAX: (609) 779-8488

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 198
TYPE: Nucleic Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
ANTI-SENSE: No
PCT-US95-10668-4

Query Match 0.9%; Score 46.4; DB 5; Length 198;
Best Local Similarity 52.6%; Pred. No. 0.0085;
Matches 101; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 1357 AAGAGTTTATCAAGCAATCAAGAGATCAAAATTGAGAGAGAAATTTGAGAGT 1416
DB 2 AACACAAAC 61
QY 1417 GACAGAAAGTTACGAGACGCGACGACAAAGTCTAAAGATATCACTAAAGATATGAA 1476
DB 62 AACACAAAC 121
QY 1477 AAGCTGCTCAAGCAGATCTATGATTCCAAATTCACATATACATGACCTGACCACTTC 1536
DB 122 AACACAAAC 181
QY 1537 GAGAAATGATG 1548
DB 182 AAGCTGATCTG 193

RESULT 17
US-08-098-327E-37
Sequence 37, Application US/08098327E
Patent No. 6270771
GENERAL INFORMATION:
APPLICANT: GUBERIN-MARCHAND, Claudine
TITLE OF INVENTION: DRUGS, PLEI
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

NAME: MCGOWAN, MALCOLM A

SOFTWARE: PatentIn Re

```

; CURRENT APPLICATION DATA:

```


APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.022;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 629 TGAAGATCAGACCAAGGAGTTGAGCTATTGAAGAAGTTGGCTTCGATATCGCAGC 668
DB 596 TAAAGAAAGTTGCAAGAACACAAAGCGATTGAAACAAGAGAGAGCTCTAAAGAAA 655
QY 689 CTCTCGACACATCAAGACATGTGGAAAGATGGAAGATTATATTAAGAATAAGA 748
DB 656 GTTGCAAGAAACAAAGCATTTAGACAAAGAGAGACTGCTAAAGAAAGTTGCAAGA 715
QY 749 AGACCATCGAAGCATTTAGAGAGCTGATGCAAGATCCAAAAGACCATTGACAAAATA 808
DB 716 ACAACAAAGCATTTAGAACAGAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAG 775
QY 809 AGATGCAACCAAGAGAGAAAGAAAGAGTTGTACAGCCCGCCAGTACGACCTGTCCA 868
DB 776 CGATTAGAACAAAGAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAGCATTTAGA 835
QY 869 TCTATTAACAACAGCTTGAAGAGCCCATTAACCTCATACGCTACTGAGAGAGCGCATAG 928
DB 836 ACAAGAGAGACTGCTTAAGAAAGTTACAAAGAGAGAGAGCATTTAGAACAGATAG 895
QY 929 ACACCTCAAGAAATGAATAATATCAAGAGACTGCTCCAGCAAGATTATGA 980
DB 896 ACTTGCTAAAGAAAGTTGCAAGAACAAAGCATTTAGAACAGAGAGA 947

RESULT 20
US-08-462-625-34
Sequence 34, Application US/08462625
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUIHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPTIOPES CAPABLE
NUMBER OF SEQUENCES: 46
CORESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-34

Query Match 0.9%; Score 46.4; DB 4; Length 988;
Best Local Similarity 45.7%; Pred. No. 0.022;
Matches 161; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 629 TGAAGATCAGACCAAGGAGTTGAGCTATTGAAGAAGTTGGCTTCGATATCGCAGC 668
DB 596 TAAAGAAAGTTGCAAGAACACAAAGCGATTGAAACAAGAGAGAGCTCTAAAGAAA 655
QY 689 CTCTCGACACATCAAGACATGTGGAAAGATGGAAGATTATATTAAGAATAAGA 748
DB 656 GTTGCAAGAAACAAAGCATTTAGACAAAGAGAGACTGCTAAAGAAAGTTGCAAGA 715
QY 749 AGACCATCGAAGCATTTAGAGAGCTGATGCAAGATCCAAAAGACCATTGACAAAATA 808
DB 716 ACAACAAAGCATTTAGAACAGAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAG 775
QY 809 AGATGCAACCAAGAGAGAAAGAAAGAGTTGTACAGCCCGCCAGTACGACCTGTCCA 868
DB 776 CGATTAGAACAAAGAGAGAGCTGCTAAAGAAAGTTGCAAGAACAAAGCATTTAGA 835
QY 869 TCTATTAACAACAGCTTGAAGAGCCCATTAACCTCATACGCTACTGAGAGAGCGCATAG 928
DB 836 ACAAGAGAGACTGCTTAAGAAAGTTACAAAGAGAGAGAGCATTTAGAACAGATAG 895
QY 929 ACACCTCAAGAAATGAATAATATCAAGAGACTGCTCCAGCAAGATTATGA 980
DB 896 ACTTGCTAAAGAAAGTTGCAAGAACAAAGCATTTAGAACAGAGAGA 947

RESULT 21
US-08-182-175A-104
Sequence 104, Application US/08182175A
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
TITLE OF INVENTION: Synthetic Storage proteins with defined structure containin
NUMBER OF SEQUENCES: 113
CORESPONDENCE ADDRESS:

Db 242 AAAGATGAAGCAATGAGAAACAAATGAAGTGTGAGAGAAATGAAGAGCTGA 301
QY 3213 AGCCGAGAT 3221
Db 302 AGAGAAAGAT 310

RESULT 23
5171843-8
; Patent No. 5171843
; APPLICANT: NUSSENZWEIG, VICTOR
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO:10:
; LENGTH: 697
5171843-10

Query Match 0.9%; Score 45.6; DB 6; Length 697;
Best Local Similarity 47.0%; Pred. No. 0.03;
Matches 141; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 1689 GGAGAAAGCACTAAAGTATTACAAAGATTCATAGTAGATCGAAAACGAGATCGAGAC 1748
Db 124 GGACATAGACAGCTGACCAACCGAGGATAGACAGATGACGCCAGCGAGAGAC 183
QY 1749 GCTTGTGAGACATTAAGAGAGATGAAGACAGTTGTTGAGAAGATTAACAAAGA 1808
Db 184 AGACGAGATGACAGCAGCAGAGAGACAGATGACAGACCAACGAGAGACAGAGCA 243
QY 1809 CGAAATTAACCAAGATGAGAGATCCCTGAGAGTCTCCGATATTGTTAAAGTCCAGTGA 1868
Db 244 GATGAGCAACACAGCTGATAGAGCAGCTGACCAACGAGCTGATAGAGCAGCTGA 303
QY 1869 GAAGTGCTCCTGATGAACAGATTGATGAACCAAGACCTCACTTCTGAAGAA 1928
Db 304 CAACGAGAGATGAGAGAGATGACAGCAGCAGAGAGATAGAGCAGCTGACAGACCA 363
QY 1929 CGTGAGATTAAACATATATATACATGTGCCGAATAGTTATAGCAGAGAGATAAGCAGA 1988
Db 364 GCAGGAGATGAGCAGATGTGACAGCCAGAGGAGATAGACAGCTGACAGCCAGCAGGA 423

RESULT 24
5171843-8
; Patent No. 5171843
; APPLICANT: NUSSENZWEIG, VICTOR
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO:8:
; LENGTH: 1137

5171843-8
Query Match 0.9%; Score 45.6; DB 6; Length 1137;
Best Local Similarity 47.0%; Pred. No. 0.04;
Matches 141; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 1689 GGAGAAAGCACTAAAGTATTACAAAGATTCATAGTAGATCGAAAACGAGATCGAGAC 1748
Db 316 GGAGACAGACAGCTGACCAACCGAGGATAGACAGATGACCAACGAGAGAGAT 375
QY 1749 GCTTGTGAGACATTAAGAGAGATGAAGACAGTTGTTGAGAAGATTAACAAAGA 1808
Db 376 AGACGAGCTGACCAACCGAGGATAGACAGATGACCAACGAGAGAGAGAGCA 435
QY 1809 CGAAATTAACCAAGATGAGAGATCCCTGAGAGTCTCCGATATTGTTAAAGTCCAGTGA 1868
Db 436 GATGAGACAGCAGCAGAGAGAGAGATGACCAACGAGAGAGAGAGAGAGATGA 495
QY 1869 GAAGTGCTCCTGATGAACAGATTTGATGAACCTCAAGACTCACTTCTGAAGAA 1928
Db 496 CAACGAGAGATGATAGAGAGCTGACCAACGAGCTGATAGAGCAGCTGACAGCA 555
QY 1929 CGTGAGATTAAACATATATATACATGTGCCGAATAGTTATAGCAGAGAGATAAGCAGA 1988
Db 556 GCAGGAGATGAGCAGATGTGACAGCCAGCAGAGATAGACAGCTGACAGCCAGCAGGA 615

RESULT 25
PCT-US95-10668-1
; Sequence 1, Application PC/TUS9510668
; GENERAL INFORMATION:
; APPLICANT: James Eberwine
; TITLE OF INVENTION: A Method of Sequencing Proteins by
; TITLE OF INVENTION: Epitope Ordering and Protein
; TITLE OF INVENTION: Restriction Mapping
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jane Massey Licata, Esq.
; STREET: 210 Lake Drive East, Suite 201
; CITY: Cherry Hill
; STATE: NJ
; COUNTRY: USA
; ZIP: 08002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
; COMPUTER: IBM 486
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10668
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/294,133
; FILING DATE: August 22, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: PENN-0137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 779-8488
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
; PCT-US95-10668-1

Query Match 0.9%; Score 45.4; DB 5; Length 198;
Best Local Similarity 53.0%; Pred. No. 0.016;


```

: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22133-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/098,327E
: FILING DATE: 24-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286
: FILING DATE: 05-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McGowan, Malcolm K.
: REGISTRATION NUMBER: 39,300
: REFERENCE/DOCKET NUMBER: 010830-045
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-6620
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 950 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 92/13884
: PUBLICATION DATE: 20-AUG-1992
:
: US-08-098-327E-32
:
: Query Match 0.9%; Score 44.8; DB 4; Length 950;
: Best Local Similarity 45.5%; Pred. No. 0.058;
: Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
:
: QY 629 TGAAGATCGAGCCCAAGCGATTGAGCGTATTGAGAAGTTGGCTTCGATATCGCAAGC 688
: DB 29 TAAAGAAAAGTTGCAAGAACCAAGCGATTGAGAACAGATAGACTGCTAAAGAAAA 88
: QY 689 CTCCTGACACATCAAGAGCAATGTGGAAAGATGGAAGATTATTAAGAAATTAAGA 748
: DB 89 GTTACAAAGAGCGCAAGAGGATTTAGAACAGAGAGACTTGTAAAGAAAAGTTGCAAGA 148
: QY 749 AGACCATCGAGAACATTTAAGAGCTGATGGAAGAAATCCAAAAAGACCATAGACAAAATA 808
: DB 149 ACAACAAAGCGATCTAGAACAGAGAGCGTCTAAAGAAAAGTTGCAAGAACACAAAGA 208
: QY 809 ACAAATGCAACCAAGAGAGAAAGAAAGAAAGTGTACACAGGCCCGCATAGACCTGTCCA 868
: DB 209 CGATTTAGAACCAAGAGAGCGTGTAAAGAAAAGTTGCAAGAACCAAGCGATTTTAGA 268
: QY 869 TCTATTAACAACAGCTTGAAGAAAGCCCATACCTCATCAGCTTACTGGAAGAGCGCATAG 928
: DB 269 ACAAGATAGACTTGTCTAAAGAAAAGTTACAGAGAGCGACGACGATTTTGAACAAAGAG 328
: QY 929 ACACCCCTCAAGAGAATGAATAATATCAAGAACTGCTCGACAGATTAATGA 980
: DB 329 ACGTGCTTAAGAAAAGTTGCAAGAACACAAAGCGATTGTAAGAACAGAGAGA 380
:
: RESULT 29
: US-08-462-625-32
: Sequence 32, Application US/08462625
: Patent No. 6319502
: GENERAL INFORMATION:
: APPLICANT: GUERIN-MARCHAND, Claudine
: APPLICANT: DRUILHE, Pierre
```

```

: TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
: TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
: OF STIMULATING THE T LYMPHOCYTES
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22133-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/462,625
: FILING DATE: 05-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/098,327
: FILING DATE: 24-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286
: FILING DATE: 05-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McGowan, Malcolm K.
: REGISTRATION NUMBER: 39,300
: REFERENCE/DOCKET NUMBER: 010830-078
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 950 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 92/13884
: PUBLICATION DATE: 20-AUG-1992
:
: US-08-462-625-32
:
: Query Match 0.9%; Score 44.8; DB 4; Length 950;
: Best Local Similarity 45.5%; Pred. No. 0.058;
: Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
:
: QY 629 TGAAGATCGAGCCCAAGCGATTGAGCGTATTGAGAAGTTGGCTTCGATATCGCAAGC 688
: DB 29 TAAAGAAAAGTTGCAAGAACCAAGCGATTGAGAACAGATAGACTGCTAAAGAAAA 88
: QY 689 CTCCTGACACATCAAGAGCAATGTGGAAAGATGGAAGATTATTAAGAAATTAAGA 748
: DB 89 GTTACAAAGAGCGCAAGAGGATTTAGAACAGAGAGACTTGTAAAGAAAAGTTGCAAGA 148
: QY 809 ACAAATGCAACCAAGAGAGAAAGAAAGAAAGTGTACACAGGCCCGCATAGACCTGTCCA 868
: DB 209 CGATTTAGAACCAAGAGAGCGTGTAAAGAAAAGTTGCAAGAACCAAGCGATTTTAGA 268
: QY 869 TCTATTAACAACAGCTTGAAGAAAGCCCATACCTCATCAGCTTACTGGAAGAGCGCATAG 928
: DB 269 ACAAGATAGACTTGTCTAAAGAAAAGTTACAGAGAGCGACGACGATTTTGAACAAAGAG 328
: QY 929 ACACCCCTCAAGAGAATGAATAATATCAAGAACTGCTCGACAGATTAATGA 980
: DB 329 ACGTGCTTAAGAAAAGTTGCAAGAACACAAAGCGATTGTAAGAACAGAGAGA 380
```



```
RESULT 30
US-08-098-327E-41
: Sequence 41, Application US/08098327E
: Patent No. 6270771
: GENERAL INFORMATION:
: APPLICANT: GUERIN-MARCHAND, Claudine
: APPLICANT: DRULHE, Pierre
: TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
: TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
: TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/098,327E
: FILING DATE: 24-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286
: FILING DATE: 05-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McGowan, Malcolm K.
: REGISTRATION NUMBER: 39,300
: REFERENCE/DOCKET NUMBER: 010830-045
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1482 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1482
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 92/13884
: PUBLICATION DATE: 20-AUG-1992
: US-08-098-327E-41

Query Match          0.9%; Score 44.8; DB 4; Length 1482;
Best Local Similarity 45.5%; Pred. No. 0.075;
Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
```

```
QY 869 TCTATACAAACAGCTTGAGAGCCCATTAACCTTCAGCGTACTGAGAGCGCATAG 928
DB 279 ACAAGATGACTCTGCTAAGAAAGAAAGTTACAGAGAGCAAGCCGTTTAGAACAAGAG 338
QY 929 ACACCTCAGAGAGATGAATAATATCAAGAAGCTGCTGACAGATTATGA 980
DB 339 ACGTCTAAGAAAGTTGCAAGAACAAAGCGATTAGAACAGAGAGA 390

RESULT 31
US-08-098-327E-45
: Sequence 45, Application US/08098327E
: Patent No. 6270771
: GENERAL INFORMATION:
: APPLICANT: GUERIN-MARCHAND, Claudine
: APPLICANT: DRULHE, Pierre
: TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
: TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/098,327E
: FILING DATE: 24-NOV-1993
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286
: FILING DATE: 05-FEB-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: McGowan, Malcolm K.
: REGISTRATION NUMBER: 39,300
: REFERENCE/DOCKET NUMBER: 010830-045
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 45:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1482 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1482
: PUBLICATION INFORMATION:
: DOCUMENT NUMBER: WO 92/13884
: PUBLICATION DATE: 20-AUG-1992
: US-08-098-327E-45

Query Match          0.9%; Score 44.8; DB 4; Length 1482;
Best Local Similarity 45.5%; Pred. No. 0.075;
Matches 160; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
```



```
Query Match      0.98; Score 44.8; DB 4; Length 1482;
```

TYPE: nucleic acid
STRANDEDNESS: single

```
Query Match      0.98; Score 44.8; DB 4; Length 1482;
```



```

? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1482
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 92/13884
? PUBLICATION DATE: 20-AUG-1992
?
? US-08-462,625-45

```

Query Match	0.98;	Score 44.8;	DB 4;	Length 1482;
Best Local Similarity	45.58;	Pred. No. 0.075;		
Matches 160; Conservative	0;	Mismatches 192;	Indels 0;	Gaps 0;

[illegible]

FILING DATE: 20-MAR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Frommer, William S.
 REGISTRATION NUMBER: 25,506
 REFERENCE/DOCKET NUMBER: 454310-2570
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 840-3333
 TELEFAX: (212) 840-0712
 TELEX: 425066 CURTMS
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2223 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-257-073-4

Query Match	0.93;	Score	44.8;	DB	1;	Length	223;
Best Local Similarity	52.78;	Pred.	No.	0.096;			
Matches	97;	Conservative	0;	Mismatches	87;	Indels	0;

Qy	718	AAAGTGGAGAGCTTATATTTAAAAAGATTAAAGAGACCATCGACAAACATTATACAGCTGTAC	777
Qy	718	AAAGTGGAGAGCTTATATTTAAAAAGATTAAAGAGACCATCGACAAACATTATACAGCTGTAC	777
Db	2033	AAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG	2092
Qy	778	GAAAGATCCAAAAAGACCATAGACAAAAAATTAAGCAATGCAACCAAGAGAGGAAAGAAAGAG	837
Db	2093	AAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG	2152
Qy	838	AAAGTTGTACAGAGGCCAGTACGACCTGTCCATCTATATCAAAACAGCTTGAAGAGACCCAT	897
Db	2153	AAGGAAGAGAAACAAGAAAGAAAGAAAGAAATAGTACAGAAAAATTTGACAACTGAAG	2212
Qy	898	AACC 901	
Db	2213	AATC 2216	

RESULT 35
 PCT-US93-07261-10
 Sequence 10, Application PC/TUS9307261
 GENERAL INFORMATION:
 TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: John H. C. Blasdale
 STREET: One Giralda Farms
 CITY: Madison
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07940-1000
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Apple Macintosh
 OPERATING SYSTEM: Macintosh 6.0.5
 SOFTWARE: Microsoft Word 5.1a
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/07261
 FILING DATE: 19930805
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/927,531
 FILING DATE: 07-AUG-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Blasdale, John H. C.
 REGISTRATION NUMBER: 31,895
 REFERENCE/DOCKET NUMBER: DX0288K
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-822-7398
 TELEFAX: 201-822-7039
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4766 base pairs
 TYPE: nucleic acid


```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 2277 base pair
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
:
US-09-098-487-2

```

Query Match	0.9%;	Score 44.2;	DB 2;	Length 2277;
Best Local Similarity	26.7%;	Pred. No. 0.14;		
Matches 185;	Conservative 102;	Mismatches 402;	Indels 3;	Gaps 1;

Oy	1625	ATCTTGGAAGAGCTACCAAGAGCTCTTAAGATATATGGAGCACTATCTCGCGAAGCATTTG	1684
Db	170	ATGTCNARMGNCGYTTAAAGARATTHACNACNTTTCARBGNITGYAARATHAAYGTNACNG	229
Oy	1685	TTTGGAAGAAGAACTAAAGTATTACAGAATCTCATAGTAAGATCGAAAACGAGATCG	1744
Db	230	TGNCNNARAARAAAYTTWNGNAAAYAAARCNABRGARABGNABGNABRAAAYGARAAYWNSGART	289
Oy	1745	AGACGCTGTTGGACACATTTAAGAGAGTGAACAACAGTGTGTTGGAGAAGAGATTACAA	1804
Db	230	GYCCNAAARABARAGCCAAABGCNABARACNABARTNGCNGATATARAARCGCMBNNTYNA	349
Oy	1805	AAGACGAAATTAACCCGATGAGAGATCCTGGAGTCCGATATTTGTTAAAGTCACG	1864
Db	350	THATHMGMAAYTWMSTTYAARTGYWSNGARGAYGAYTTNABACNGTNTTGYGCNART	409
Oy	1865	TGCGAAGAGTGCTCCTCATATGACACAGATTGATTAAGTCAAGAAGACTCAACTCATTTG	1924
Db	410	TYGNGCGCTGTTTGARGTAAATATTCNMGNAARCNAGAYGNNABARATGMNGNTTYG	469
Oy	1925	AGAAGCTGGAGTTTAAACATTAATATACATGTCGCCAATAGTTATTAAGCGAGAAATAAGC	1984
Db	470	GNTTYGTNCATTTTAAATAAYTTWTNGARBCCNGNABRCNTTNAABGNATGAAYATGA	529
Oy	1985	AGCAACCATTAACCTA---TGTACTCAAGAAAGAGATAGACAACACTGAAAGTGTCA	2041
Db	530	ARGARATHAAGNMGNACNGTNGCNGTNGATYGGCNGTNGCNAABRGATTAARTYYAARG	589
Oy	2042	TGCCCAAAGTCGAGAGGCTCATCAAGACAGAGAAGAAACATTAAACTGAAGACAGT	2101
Db	590	AYACNCAKMSNGTWMNGCNATHGNCNARBARABARBSNCAITGARBMNABARCAIYCARGARW	649
Oy	2102	CAGATTAACCTCCGACCTTCCACAGAAAGAGAGATAACCGACAGGCTACCCAAAGCCCG	2161
Db	650	SNGTNAAARAAARAAAGMNGMAGNARGARBARAYTTGARBARGARBARAAYGAYGAYG	709
Oy	2162	GACAACAGGCGCGTTTCAGCTCTGGAAGGGCATAGCGTCAAGGCTCAAGCAAGAGACGA	2221
Db	710	ATGATGATGATGATGARGARGATGNGCTNTTYATGATYGARGATYGARGARGARAAAYA	769
Oy	2222	AGCAGGACACAGCTTCAGTGGCCAGTGGCCGTTCCAGAGGCTAAAGCTCAAGTGCCTACAG	2281
Db	770	THGARMSNAAAGTACNACNABRCNCTNCATRATHCARAABRMGNCNGTNAABRMGNCNGNC	829
Oy	2282	CACCAAGCTCCTGTGATTAACAACAAGCGAGAAAT	2313
Db	830	CNGCNAABRWMSNMGATCAATWNSGARCABAY	861

RESULT 39
US-08-559-8968-1
Sequence 1, Application US/08559896B
Patent No. 6310046
GENERAL INFORMATION:
APPLICANT: Patrick E. Duffy
APPLICANT: Christian F. Ockenhouse
TITLE OF INVENTION: SEQUESTIRIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: John Moran
STREET: USA MRC - MCMR-JA

CITY: FORT DETRICK, FREDERICK
 STATE: MARYLAND
 COUNTRY: USA
 ZIP: 21702-5012
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Apple Macintosh
 OPERATING SYSTEM: Macintosh 7.5
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/559,896B
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Moran, John
 REGISTRATION NUMBER: 26,313
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 619-2065
 TELEFAX: (301) 619-7714
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1956 base pairs
 TYPE: Nucleic acid
 STRANDEDNESS: Double
 TOPOLOGY: Linear

US-08-559-896B-1

Query Match	0.98;	Score 43.6;	DB 4;	Length 1956;
Best Local Similarity	42.9%;	Pred. No. 0.19;		
Matches 217;	Conservative	0;	Mismatches 289;	Indels 0;
				Gaps 0.

QY	1360	AAGTTTTCACGAAATTC	AAAGGGAAGATCA	CAAAATTCG	AACACACAA	CAAAATTCG	AGAGTAC	1419
Db	1132	AATTTAAGCGAAATTC	AAAAGGTAATTTAT	TATTCATAG	AAATTTAAAAAT	TGAATCTTAC	1120	
QY	1420	AAGAAAAGTTACG	AGACGCCACG	CAAAAGCTT	AAACGATATC	CTAAGAGATG	AAAAA	1479
Db	1192	CAAAAAAATTAAG	AAAGAAATGT	AAATGTATTTAT	TATTAATAC	GACGACAAATAT	AGTAT	1251
QY	1480	CTGCTGACGAGAT	CTATGATTC	CAAAATTC	AAACATAT	AACATGACCTG	ACCACTTGAG	1539
Db	1252	AATATATATATAT	ATATATATATAG	ATGTCGAATATTT	AAATAAAT	AAGACACAAATAT	1311	
QY	1540	AAAAATGATGG	AAAAACGGTACT	CTTCAAGAGTG	AGAAATC	CAACACCATTAAT	ACTTT	1599
Db	1312	AATTTATGAAAA	ATGTAGATG	AGTTGAATTTG	TATGCAAAATTT	TGACACAGATTA	AAAGA	1371
QY	1600	GCATCTATGAG	AATTCCTAG	CAATATCTTG	AGAACCTC	CAACAAAGCTTT	TAAGTATG	1659
Db	1372	GCCAAATAG	AGATATTTAT	TAGCTTTATTT	TAACAAAGAAATTT	AAAAAAGACAAAAAT	GT	1431
QY	1660	GAGGACATTTCT	CGCGGAATTTG	TGTGGAGAA	GAACCTAA	AGTATTTACA	GAATCTC	1719
Db	1432	AATGTTTCC	ATATAGGATTTT	TAAATTC	AAAAGTGA	AAAAAGATTA	CAACACCAATTT	1491
QY	1720	ATAAGTAAAT	CGAAAAACG	AGATCG	AGCGCTTTG	GAGACATTT	PAGAGAGATGA	1779
Db	1492	CAACATTAAG	AGAAATCTA	AGTAGATTTG	TGCAGGAAAAAT	TATTCAGTTT	ATTCAMAG	1551
QY	1780	CAGTTGTTT	GAGAGAGATTT	ACAAAAAG	ACGAAATTA	AAACCGAT	TAGAGAGATCTCGAG	1839
Db	1552	GATATATTA	AAAAAATTA	AGGCCAAAA	AGGATTA	ACACTGA	ATATTTAGTAATTA	1611
QY	1840	GCTCCGATAT	TGTTAAGTCC	AGT	1865			
Db	1612	ATAACAAATAT	TGATAT	AAAAATGT	1637			

RESULT 40

	Query Match	0.9%	Score 43	DB 1	Length 1241
	Best Local Similarity	44.7%	Pred. No. 0.2		
	Matches	215	Conservative	0	Mismatches 260
					Indels 6
					Gaps 1
QY	2972	AGAACTGGAGAGACATCATTAAGCTCGAAGAAACACTGCACTGAGCTTCGACCTGT	3031		
DB	88	AGAAAGGAAAGAGAGGGGAGAGCTTACCGCCACCGAAGAGGCAAGATGAACAACCTCTCG	147		
QY	3032	ACAACAACTACAAACTGAAACTGGAGAGACTCTTCGACAGAGAAGACAGTCGGCAACT	3091		

RESULT 41
 : US-08-471-033-42
 : Sequence 42, Application US/08471033
 : Patent No. 3770696
 : GENERAL INFORMATION:
 : APPLICANT: Warren, Gregory W
 : APPLICANT: Kozielec, Michael G
 : APPLICANT: Mullins, Martha A
 : APPLICANT: Nye, Gordon J
 : APPLICANT: Carr, Brian
 : APPLICANT: Desai, Nalin M
 : APPLICANT: Kostichka, N. Kristy
 : APPLICANT: Duck, Nicholas B
 : APPLICANT: Estruch, Juan J
 : TITLE OF INVENTION: NO. 5770696el Pesticidal Proteins and Strains
 : NUMBER OF SEQUENCES: 50
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: CIBA-GEIGY Corporation
 : STREET: 7 Skyline Drive
 : CITY: Hawthorne
 : STATE: NY
 : COUNTRY: USA
 : ZIP: 10532
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30B
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/471,033
 : FILING DATE:
 : CLASSIFICATION: 530
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/314,594
 : FILING DATE: 09-SEP-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/218,018
 : FILING DATE: 23-MAR-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/037,057
 : FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: P-40,403
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1241 base pairs
 type: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "Synthetic DNA"
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..1238
 OTHER INFORMATION: /note= "Maize optimized DNA
 OTHER INFORMATION: sequence encoding Vir2A(a) with the Bacillus secretion signal
 OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
 contained in pCIB528"

Query Match	0.98	Score 43	DB 1	Length 1241
Best Local Similarity	44.78	Pred. NO. 0.2		
Matches 215	Conservative	0	Mismatches 260	Indels 6
				Gaps 1

OY	2972	AGAACTCGAAGAGACATCAATTAACCTGAAAGACACTGCAACACTTCGACCT	3031
Db	88	AGAAAGGAAGAGAGTGGAAACCTTACCGCCACCGAAGGGCAAGATGAACAACTTCCTGG	147
OY	3032	ACAACAAGTACAAACTGAACTGAGAGACTTTCGACAGAAGAAGACATCGCGCAAGT	3091
Db	148	ACAACAAGAAGACGACATTCAGACCCAACTACAGAGAGATCACTTCAGCATATGCCGCGACCT	207
OY	3092	ATAAGATGACAGATCAACGAATGTAGCTCGTCCAAAGAGAGAGCTTGA-----AAGCAAC	3145
Db	208	TCGAGAGACGAGATCAAGGACCTGAGAGATGCAAGAATGTTTCGACAAGACCAACTCGA	267
OY	3146	TCAACTCACTGAACAATCCGAAACACGTACTCGAAGCTTCTCAAGTGTCTTCAACAAGA	3205
Db	268	GCAACACACATCATCACCCTACCAAGACGTGGAGCCCAACCTCGCTTCAACAAGAGCC	327
OY	3206	AGAAAGGAAGCGAGATGGCGGAGACAGAGAACCTCTGGAGAACCAAGATTCTTCACA	3265
Db	328	TGACCCGAGGGCAACACCATATACAGGACGCCATGGCCCAAGTTCAAGAGACATCTCGG	387
OY	3266	AACACTACAAAGGCGCTCGTCAAGTATTATTAATGGAGATCTTCTCCTCGAAGACTCTCT	3325
Db	388	ACCCGCGATCAAGTTCTGACAGCTACTCTGGACACCACTGACCGCCGACAGGTGAGACA	447
OY	3326	CCGAGGAGACATCCAGACCGGAGGATTAACCTACGCCACGCTCGAGAACTTCAAGGTCTGT	3385
Db	448	GCAAGGAGCGCGGTATCTCTGAAGGTGACCGTCCCAAGCGGCAAGGCAACACACCCCA	507
OY	3386	CTAAGCTCGAAGGCAAGCTAAAGACAACCTGCAACCTGGAGAAGAAGAAGCTCAACTAC	3445
Db	508	CCAAAGCGCGGGGTATCTCTGAACAACAAGGAGTACAAAGATGCTGATCGACAAAGGCTACA	567
OY	3446	T 3446	
Db	568	T 568	

RESULT 42
US-08-471-044-39
; Sequence 39, Application US/08471044
; Patent NO. 5440868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziele, Michael G

```

1  APPLICANT: Mullins, Martha A
2  APPLICANT: Nye, Gordon J
3  APPLICANT: Carr, Brian
4  APPLICANT: Desai, Nalin M
5  APPLICANT: Kostichka, N. Kristy
6  APPLICANT: Duck, Nicholas B
7  APPLICANT: Estruch, Juan J
8  TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
9  NUMBER OF SEQUENCES: 50
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: CIBA-GEIGY Corporation
12 STREET: 7 Skyline Drive
13 CITY: Hawthorne
14 STATE: NY
15 COUNTRY: USA
16 ZIP: 10532
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patentln Release #1.0, Version #1.30B
23
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/471,044
26 FILING DATE: 06-JUN-1995
27 CLASSIFICATION: 800
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 08/463,483
30 FILING DATE: 05-JUN-1995
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER: US 08/314,594
33 FILING DATE: 09-SEP-1994
34 PRIOR APPLICATION DATA:
35 APPLICATION NUMBER: US 08/218,018
36 FILING DATE: 23-MAR-1994
37 PRIOR APPLICATION DATA:
38 APPLICATION NUMBER: US 08/037,057
39 FILING DATE: 25-MAR-1993
40 ATTORNEY/AGENT INFORMATION:
41 NAME: Pace, Gary M.
42 REGISTRATION NUMBER: 40,403
43 REFERENCE/DOCKET NUMBER: CGC 1695/CTP3/DIV6 - SOLV3
44 TELECOMMUNICATION INFORMATION:
45 TELEPHONE: 919-541-8582
46 TELEFAX: 919-541-8689
47 INFORMATION FOR SEQ ID NO: 39:
48 SEQUENCE CHARACTERISTICS:
49 LENGTH: 1241 base pairs
50 TYPE: nucleic acid
51 STRANDEDNESS: single
52 TOPOLOGY: linear
53 MOLECULE TYPE: /desc = "nucleic acid
54 DESCRIPTION: /desc = "Synthetic DNA"
55 HYPOTHETICAL: NO
56 FEATURE:
57 NAME/KEY: CDS
58 LOCATION: 9..1238
59 OTHER INFORMATION:
60 OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion sig
61 OTHER INFORMATION: removed as contained in PCIB557"
62
63 US-08-471-044-39

```

Query Match	0.98	Score 43	DB 2	Length 1241
Best Local Similarity	44.78	Pred. NO. 0.2		
Matches 215, Conservative	0	Mismatches 260	Indels 6	Gaps 1

[illegible]

QY 3092 ATAGATCAGATCAAGAAGTTGACTCTGCTCAAGAGACGCTGA-----AGCAAAAC 3145
DB 208 TCGAGAGAGAGATCAAGACCTGGAAGAGATGACAGATGTTCCACAAGACCACTGA 267
QY 3146 TCAATCTACTGACATCCGAAACACGTAAGTCTGAGAACTTCTGAGTTCTTCAACAGA 3205
DB 268 GCAAGCAGATCTACCTACCAAGAACGTGGACCCACCCACCTCGCTTCAACAAGAGCC 327
QY 3206 AGAAGAGCCGAGATGCGCCGAGACAGAGAACATCTGAGAACCAACCAATTTCTTCA 3265
DB 328 TGACCGAGAGGCAACACATCAACAGCAGACGCGATGCGCCAGTTCAAGGAGCTTGG 387
QY 3266 AACACTACAAAGCCCTCGTCAAGTATTAATGAGGAGTCTTCTCTGAGAGCTCTCT 3325
DB 388 ACCGGACATCAAGTTCCAGACCTGACCTGAGACCCACCTGAGCCGCGAGGAGAGA 447
QY 3326 CCGAGAGAGATCCAGACCGGAGTAATCTAGCCAGCTCGAGAACTTCAAGTCTCT 3385
DB 448 GCAAGGAGCGCTGATCTGTAAGGTGACCGTCCCGAGGCGCAAGGCGACACCCCA 507
QY 3386 CTAAGCTGGAAGGCAAGCTGAAGCAACCTGAGCTGAGAGAGAAAGAGCTGAGTACC 3445
DB 508 CCAAGGCGGCGTGAATCTGAAACAACGAGTACAGATGCTGATCGACAAGGCTACA 567
QY 3446 T 3446
DB 568 T 568

RESULT 43
US-08-471-044-42
; Sequence 42, Application US/08471044
; Patent No. 5840868

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estuch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993

ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion sig
OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
OTHER INFORMATION: contained in pCB5528"
US-08-471-044-42

Query Match 0.9% Score 43; DB 2; Length 1241;

Best Local Similarity 44.7%; Pred 0.0.2; Mismatches 260; Indels 6; Gaps 1;

Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 2972 AGAAGCTGGAAGAGACATCAATTAAGCTGGAAGAGACACTGCACTGACCTGT 3031
DB 88 ACAAGAGAGAGAGTGAAGAGCTTACCGCCACGAGAGGCAAGATGAACCACTTCTGG 147
QY 3032 ACAACAAGTACAAGTGAAGTGAAGTGAAGAGACTTGTGACAAAGAGAGAGTGGCACT 3091
DB 148 ACAACAAGAGAGACATCAAGACCAACTACAAAGAGATCACTTGAAGATGCGGAGCT 207
QY 3092 ATAGATGAGATCAAGAAGTGAAGTCTGCTCAAGGAGAGCTTGA-----AGCAAAAC 3145
DB 208 TCGAGAGAGAGATCAAGAGACCTGGAAGAGATGACAAAGATGTTCCACAAGACCACTGA 267
QY 3146 TCAATCTACTGACATCCGAAACACGTAAGTCTGAGAACTTCTGAGTTCTTCAACAGA 3205
DB 268 GCAAGCAGATCTACCTACCAAGAACGTGGACCCACCCACCTCGCTTCAACAAGAGCC 327
QY 3206 AGAAGAGCCGAGATGCGCCGAGACAGAGAACTCTGAGAACTTCTGAGTTCTTCAACAGA 3265
DB 328 TGACCGAGAGGCAACACATCAACAGCAGACGCGATGCGCCAGTTCAAGGAGCTTGG 387
QY 3266 AACACTACAAAGCCCTCGTCAAGTATTAATGAGGAGTCTTCTCTGAGAGCTCTCT 3325
DB 388 ACCGGACATCAAGTTCCAGACCTGACCTGAGACCCACCTGAGCCGCGAGGAGAGA 447
QY 3326 CCGAGAGAGATCCAGACCGGAGTAATCTAGCCAGCTCGAGAACTTCAAGTCTCT 3385
DB 448 GCAAGGAGCGCTGATCTGTAAGGTGAGACCGTCCCGAGGCGCAAGGCGACACCCCA 507
QY 3386 CTAAGCTGGAAGGCAAGCTGGAAGGACAACTGAGAGAAAGAGAGTCTGAGTACC 3445
DB 508 CCAAGGCGGCGTGAATCTGAAACAACGAGTACAGATGCTGATCGACAAGGCTACA 567
QY 3446 T 3446
DB 568 T 568

RESULT 44

US-08-463-483A-39

; Sequence 39, Application US/08463483A

; Patent No. 5849870

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G

APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spurrill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 1241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 9..1238
OTHER INFORMATION: /note= "maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP2(a) with the Bacillus secretion signal
OTHER INFORMATION: removed as contained in pcIB5527
US-08-463-483A-39
Query Match 0.9%; Score 43; DB 2; Length 1241;
Best Local Similarity 44.7%; Pred. No. 0.2;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 3146 TCAACTCACTGAACATTCGAAACAGTACTGCAAGACTTCTCAGTCTTCTTCACACAGA 3205
DB 268 GCACAGCATCATCACCCTACAAAGACGTGGAGCCACCAACCATCGGCTTCAACAAGAGCC 327
QY 3206 AGAAGAGAGCCGAGATCGCCGAGACAGACACTCTGGAGAACACCAAGATTCTTCTCA 3265
DB 328 TGACCCAGAGGCAACACCATTCACAGGACGCCATGGCCCCCAAGGAGGACACTTCTCT 387
QY 3266 AACACTACAAAGGCGTCTGCTCAATATTATATGCGAGTCTTCTCTCGAAGACTCTCT 3325
DB 388 ACCGACATCAAGTTGACAGAGTACCTGACACCCACCTGACGCCGACAGAGTAGCA 447
QY 3326 CCGAGAGAGCATCCAGACGAGATTAACCTGACGACCTCGAGAACTTCAAGTCTCTG 3385
DB 448 GCAAGGAGCGCGTATCTGTAAGGTGACCGTCCCGAGCGGCAAGGAGGACAGACCCCA 507
QY 3386 CTAAGCTCGAAGGCAAGCTGTAAGAGACACTGAACCTGGAGAGAAAGAAAGTCACTAC 3445
DB 508 CCAAGGCGCGCGTATCTGTAAGGTGACCGTCCCGAGCGGCAAGGAGGACAGACGCTTCA 567
QY 3446 T 3446
DB 568 T 568
RESULT 45
US-08-463-483A-42
Sequence 42, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spurrill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615


```

DB 268 GCAACGACATCATCACCTACAAAGACGTGGAGCCCAACACCATCGCTTCAACAGAGCC 327
QY 3206 AGAAGAGAGCGGATCGCGGAGACAGACACTCTGGAGAACACCAATCTTCTCA 3265
DB 328 TGACCCAGGACCAACCATCATCAACAGCGCCCATGGCCCGTTCAAGAGCACTCTGG 387
QY 3266 AACCTACAAAGGCGCTGCTCAAGTATTAATGCGAGTCTTCTCTGAAGACTCTCT 3325
DB 388 ACCGACATCAAGTTGACAGCTACCTGGACACCACTGACCGCCGACAGAGTGAACA 447
QY 3326 CCGAGAGACATCCAGACCGAGGATTAACCCGACGCTGAGAACTTCAAGTCTGT 3385
DB 448 GCAGAGAGCGCGTGTATCTGAAGGTACCGTCCCGACGCGCAAGGCGCACCCCA 507
QY 3386 CTAAAGCTCGAAGGCAAGTGAAGACACCTGAACCTGAGAGAAAGAACTGACCTACC 3445
DB 508 CCAAGGCCGCGGTGATCTGTAACACAGCGAGTACAAAGATGCTGATGACACAGGCTACA 567
QY 3446 T 3446
DB 568 T 568

```

```

RESULT 47
US-08-471-046A-42
; Sequence 42, Application US/08471046A
; Patent No. 5866326
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5866326artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,046A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ. ID NO.: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1241 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9...1238
; OTHER INFORMATION:
; OTHER INFORMATION: sequence encoding VIP2a(a) with the Bacillus secretion sig
; OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
; OTHER INFORMATION: contained in pCIB5358"
US-08-471-046A-42

```

```

Query Match 0.9%; Score 43; DB 2; Length 1241;
Best Local Similarity 44.7%; Pred. No. 0.2;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

```

```

QY 2972 AGAAGCTGAAGAGGACATTAATAGCTGAAGAAGACACTGCACTGAGCTTGACCTGT 3031
DB 88 AGAAGGAGAGAGGTGAGGAACTTACCGCCACGGAAGGCGCAAGATGAAACACTTCTCTG 147
QY 3032 ACANCAAGTCAACATGCAAACTGGAGAGACTTTCGACAAAGAAAGAGAGCTGGGCAAGT 3091
DB 148 ACAACAGAGAGGACATCAATCAACCACTACAAAGAGATCACTTTCAGCATAGCCGGAGCT 207
QY 3092 ATAAAGTCAAGATCAAGAAATGACTCTGCTCAAGAGAGAGCTTGA-----AAGCAAA 3145
DB 208 TCAGAGAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 267
QY 3146 TCACATCACTGAACATTCGAAACAGCTACTGCAAGAGAGAGAGAGAGAGAGAGAGAG 3205
DB 268 GCAACGACATCATCACTCAACAAAGAGTGGAGCCACACCATTCGCTTCAACAAAGAGCC 327
QY 3206 AGAAGAGAGCGGATCGCCGAGACAGAGAACTCTGGAGAAACCAACATTCCTTCTCA 3265
DB 328 TGACCGAGGCGCAACCATTAACAGCGCGCATGGCCCGTTCAAGAGCACTTCTCTGG 387
QY 3266 AACCTACAAAGGCGCTGCTCAAGTATTAATGCGAGTCTTCTCTGAAGACTCTCT 3325
DB 388 ACCGACATCAAGTTGACAGCTACCTGGACACCACTGACCGCCGACAGAGTGAACA 447
QY 3326 CCGAGAGAGACATCCAGACCGAGGATTAACCGCCAGCTGAGAACTTCAAGTCTGT 3385
DB 448 GCAGAGAGCGCGTGTATCTGAAGGTACCGTCCCGACGCGCAAGGCGCACCCCA 507
QY 3386 CTAAAGCTCGAAGGCAAGCTTAAGAGCAACCTGAACCTGAGAGAAAGAAAGAACTTCA 3445
DB 508 CCAAGGCCGCGGTGATCTGTAACAAACAGCGAGTACAAAGATGCTGATGACACAGGCTACA 567
QY 3446 T 3446
DB 568 T 568

```

```

RESULT 48
US-08-470-566B-39
; Sequence 39, Application US/08470566B
; Patent No. 5872212
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M

```



```

1  APPLICANT: Kostichka, N. Kristy
2  APPLICANT: Duck, Nicholas B
3  APPLICANT: Estruch, Juan J
4  TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
5  NUMBER OF SEQUENCES: 52
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: No. 5872212artis Corporation
8  STREET: 3054 Cornwallis Road
9  CITY: Research Triangle Park
10 STATE: NC
11 COUNTRY: USA
12 ZIP: 27709
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30B
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/470,566B
20 FILING DATE: 06-JUN-1995
21 CLASSIFICATION: 530
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/463,483
24 FILING DATE: 05-JUN-1995
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/314,594
27 FILING DATE: 09-SEP-1994
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 08/218,018
30 FILING DATE: 23-MAR-1994
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER: US 08/037,057
33 FILING DATE: 25-MAR-1993
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Meigs, J. Timothy
36 REGISTRATION NUMBER: 38,241
37 REFERENCE/DOCKET NUMBER: GCCL695/CIP3/DIV4 - SOLV4
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: 919-541-8587
40 TELEFAX: 919-541-8689
41 INFORMATION FOR SEQ ID NO: 39:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 1241 base pairs
44 TYPE: nucleic acid
45 STRANDEDNESS: single
46 TOPOLOGY: linear
47 MOLECULE TYPE: other nucleic acid
48 DESCRIPTION: /desc = "Synthetic DNA"
49 HYPOTHETICAL: NO
50 FEATURE:
51 NAME/KEY: CDS
52 LOCATION: 9..1238
53 OTHER INFORMATION: /note= "Maize optimized DNA
54 OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
55 OTHER INFORMATION: removed as contained in PCIB5527"
56 US-08-470-566B-39

```

OY	3146	TCAACTCACTGAAGCAATTCGGAACACAGTACTGAGAACTCTCAGTGTCTTCAACAGAGA	3205
Db	268	GCAACAGCATATCATACCTTACAAAGCTGGAGGCCACACCTGSGCTTCAACAAGAGCC	327
OY	3206	AGAAGGAAGCCGAGATGCGCGAGACAGAGAACCTCTGGAGAAACCAAGATTCTTCTCA	3265
Db	328	TGACCCGAGGGCACAACCATTAACAAGGAGCCCATGGCCCATTTCAAGNACAGTCTCTGG	387
OY	3266	AACACTCAAAAGGCGCTCGTCACAGTATTATTAATGGCAGTCTTCTCCTGGAAGACTCTCT	3325
Db	388	ACCCGCAATCATAGTTGACAGCTACCTGGACACCCACTGACGCCCCACGACAGTGTAGACA	447
OY	3326	CCGAGGAGAGCATTCAGACCGAGGATTACTAGCCACAGCCTCGAGAACTTCAAGGTCTGT	3385
Db	448	GCAAGGAGCGGTGATCTGTGAAGGTGACCGTCCCGCAGCGGCAAGGGCAACACACCCCA	507
OY	3386	CTAAGCTCGAAGGCACGCTTAAGGACACACTGCAACTGGAGAGAAGAGCTCAGCTACC	3445
Db	508	CCAAGGCCCGCGCTGATCTCTGTACACAGCAGAGTACAAAGATGCTGATCGACAAACGCGTACA	567
OY	3446	T 3446	
Db	568	T 568	

US-08-470-5668-42
Sequence 42, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5872212artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatencIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4

Db 268 GCACAGCATATACCTACAGAAAGACGTGAGCCACACCATCGGCTTCAACAAGAGCC 327
Qy 3206 AGAAGAAAGCCGAGATGCCCGAGACAGAACTCTGGAGAACCAAGATTCTTCA 3265
Db 328 TGACCGAGGGCAACCCATCAACACGCGCATGCGCCAGTCAAGGACAGTTCTGG 387
Qy 3266 AACACTACAAAGGCGCTGCTCAAGTATATTAATGGGAGTCTTCTCTGTAAGACTCT 3325
Db 388 ACCGGACATCAAGTTTCACAGCTTACCTGACACCCACCTGACCGCCAGAGGTGACA 447
Qy 3326 CCGAGAGAGCATCCAGACCGAGATACTACGCGCTGAGAACTTCAAGTCTGT 3385
Db 448 GCAAGGACCGGTGATCTGTAAGGTGACCGTCCCGAGCGGCAAGGCGACACACCCCA 507
Qy 3386 CTAAGCTGGAAGGCAAGCTGGAAGCAACCTGACCTGGAGAGAAGAGTCACTACC 3445
Db 508 CCAAGGCGCGGTGATCTGTAACAACACGAGTACAGATGCTGTGACAAACGGCTACA 567
Qy 3446 T 3446
Db 568 T 568

RESULT 51
: Sequence 42, Application US/08469334
: Patent No. 5990383
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziele, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,334
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/463,483
: FILING DATE:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Spruill, W. Muray
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8615

TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1241 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "Synthetic DNA"
: HYPOTHETICAL: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 9..1238
: OTHER INFORMATION:
: OTHER INFORMATION: removed and the eukaryotic secretion signal inserted as
: OTHER INFORMATION: contained in pcIB5528
US-08-469-334-42

Query Match 0.9%; Score 43; DB 2; Length 1241;
Best Local Similarity 44.7%; Pred. No. 0.2;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

Qy 2972 AGAAGCTGGAAGGAGCATCAATAGCTGAAGAGACACCTGCAACTGAGTTCGACTGT 3031
Db 88 AGAAGAGAGAGAGTGAAGTACCGCCACCGAAGAGGCAAGTGAACAACCTTCTGG 147
Qy 3032 ACACAGTACAACTGAACTGGAGAGACTTTCGACAGAGAAGACAGTGGCAAGT 3091
Db 148 ACAACAGAGAGACATCAAGACCAACTACAGAGAGATCCTTCGATAGCGGCGACT 207
Qy 3092 ATAAGATGAGATCAAGAAAGTTGACTGCTCAAGGAGCAGCTGA-----AAGCAAG 3145
Db 208 TCGAGGACGAGATCAAGACCTGGAAGGAGATGACAAATGTTGCACAAGACCACTGA 267
Qy 3146 TCAACTCACTGAAACATCCGAACACGTACTGACAGAACTTCTCACTGTTCTTCAACA 3205
Db 268 GCACAGCATATACACTACAAAGAGGTGAGGCCACACCATCGGCTTCAACAAGAGCC 327
Qy 3206 AGAAGAAAGCCGAGATGCCCGAGACAGAAACACTCTGGAGAACCAAGATTCTTCA 3265
Db 328 TGACCGAGGGCAACCACTCAACACGACGCGATGCGCCAGTTCGAAGAGAGTTCGG 387
Qy 3266 AACACTACAAAGGCGCTGCTCAAGTATTAATGGGAGTCTTCTCTGTAAGACTCT 3325
Db 388 ACCGGACATCAAGTTTCACAGCTTACCTGAGACACCACCTGACCGCCAGAGGTGACA 447
Qy 3326 CCGAGAGAGCATCCAGACCGAGATACTACGCGCTGAGAACTTCAAGTCTGT 3385
Db 448 GCAAGGACCGGTGATCTGTAAGGTGACCGTCCCGAGCGGCAAGGCGACACACCCCA 507
Qy 3386 CTAAGCTGGAAGGCAAGCTGGAAGCAACCTGACCTGGAGAGAAGAGTCACTACC 3445
Db 508 CCAAGGCGCGGTGATCTGTAACAACACGAGTACAGATGCTGTGACAAACGGCTACA 567
Qy 3446 T 3446
Db 568 T 568

RESULT 52
: Sequence 39, Application US/09300529
: Patent No. 6066783
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziele, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B


```

1  APPLICANT: Estruch, Juan J
2  TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
3  NUMBER OF SEQUENCES: 50
4  CORRESPONDENCE ADDRESSES:
5  ADDRESSEE: No. 6066783artis Corporation
6  STREET: 3054 Cornwallis Road
7  CITY: Research Triangle Park
8  STATE: NC
9  COUNTRY: USA
10 ZIP: 27709
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: Patent In Release #1.0, Version #1.308
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/09/300,529
19 FILING DATE: TBA
20 CLASSIFICATION:
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 08/469,334
23 FILING DATE: 06-JUN-1995
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 08/463,483
26 FILING DATE: 05-JUN-1995
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 08/314,594
29 FILING DATE: 09-SEP-1994
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/218,018
32 FILING DATE: 23-MAR-1994
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/037,057
35 FILING DATE: 25-MAR-1993
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Meigs, J. Timothy
38 REGISTRATION NUMBER: 38,241
39 REFERENCE/DOCKET NUMBER: S-19506L
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: 919-541-8689
42 TELEFAX: 919-541-8689
43 INFORMATION FOR SEQ ID NO: 39:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 1241 base pairs
46 TYPE: nucleic acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49 MOLECULE TYPE: other nucleic acid
50 DESCRIPTION: /desc = "Synthetic DNA"
51 HYPOTHETICAL: NO
52 FEATURE:
53 NAME/KEY: CDS
54 LOCATION: 9..1238
55 OTHER INFORMATION: /note= "Maize optimized DNA
56 OTHER INFORMATION: sequence encoding VIP2A(a) with the Bacillus secretion signal
57 removed as contained in pcIB5527"
58
59 US-09-300-529-39

```

Query Match	Similarity	Score	DB	Length
Best	Local	44.7%	Pred. No.	0.2;
Matches	215;	Conservative	0;	Mismatches
			260;	Indels
				Gaps
				1
Qy	2972	AGAACTGTGAAGAGACATCATTAAGCTTAAAGAGACACTGTGAACCTTGCACCTT	3031	
Db	88	AGAAAGAGAGAGAGATGTGGAAAGCTTACCAGCCAGAGAGAGGCAAGATGAAACACTTCTCG	147	
Qy	3032	ACAACAAGTACAACTGAAACTGGAGAGACTCTTGCACAAAGAGAAGACAGTGGCAAGT	3091	
Db	148	ACAACAAGAGACATCAACAGACCACTACCAAGAGATGATCATTGACGATTAAGCCGAGAGCT	207	
Qy	3092	ATTAAGTGCAGATCAAGAAGTTGACTCTGCTCAAGGAGCAGCTTGA-----AAGCAAAC	3145	
Db	208	TTCAGAGACAGATGTCAAGGACCTTGAAGAGAAATGACAAAGATGTTTGCAGAACCAACTGA	267	

```

OY      3146   TCACACTCAGAACAAATCCGAAAACCAGTACGCAGAATTCTTCAGTGTTCTCACACAGA    3202
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
Db       268   GCAACAGCATCATCTACCTACAAAGACGTGGAGGCCCAACCACTCGGCTTTCAAACAAGACC    327
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
OY      3206   AGAAGGAAGCCGAGATATGCCCGAGACAGAGAACACTCTGGAGAACAACAAGATTCTTCA    3265
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
Db       328   TGACCGAGGGGCAACACCATCATCAACAGGAGAGCCCATGGCCCCAGTTCAAAGGAGCAGTTCTGG    387
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
OY      3266   AACACTACAAAGCCCTGTGAAGTAATATATATATGCGCAGTCTTCTCTCTGANAAGACTCTCT    3325
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
Db       388   ACCCGGACATCAATTCTTGACAGAGTACTTGACACCCACCACTGACGCCCAAGCAAGGTGAGCA    447
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
OY      3326   CCGAGAGAGAGCATCCAGACCGAGATAACTACGCCAGCCTTCAGAACTTCAGATGCTCTGT    3385
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
Db       448   GCAAGGAGCCGGCTGATCTCTTAAGBTGACBCTCCCAAGCGGCAAGGGCAGACACCAACCCCA    507
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
OY      3386   CTAACTTCGAAGGCAAGCTGAGGACAACTGAACCTGAGAGAGAAGAAGTCACTACCTACG    3445
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
Db       508   CCAAGCGCGGCGGTGATCTCTGACAAACAGGAGTAGCAAGATGCTGATCGACACAGCGCTACA    567
          ||||| |-----+-----+-----+-----+-----+-----+-----+-----+
OY      3446   T 3446
          +-----+-----+-----+-----+-----+-----+-----+-----+
Db       568   T 568
          +-----+-----+-----+-----+-----+-----+-----+-----+

RESULT 53
US-09-300-529-42
; Sequence 42, Application US/09300529
; Patent No. 6066783
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory M
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Cairr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostlichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6066783artlis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/300,529
; FILING DATE: TBA
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/469,334
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
```


Query Match	0.98;	Score 43;	DB 3;	Length 1241;
Best Local Similarity	44.7%;	Pred. No. 0.2;		
Matches 215;	Conservative	0;	Mismatches 260;	Indels 6;
				Gaps 1;

QY	2972	AGAAAGCTGGAGAGGAGCATCATATTAAGCTGTGAAGAAAGCACTGCACTGGAGCTTGACACGT	3013
Db	88	AGAAAGGAAAGAGAGTGGAGACTTACCGCCACCGAGAAAGGGCAAGATGAAACAATCTTCGG	147
QY	3032	ACAAACAAGTACAAACTGAAACTGGAGAGACTCTTCGCACAAGAAAGAACAGTCGGCAAGT	30913
Db	148	ACAAACAAGAGACATCAAGACCAACTACAGAGATCACCTTCAGCATAGCGGGACGCT	207
QY	3092	ATAAGATGCAGATCAAGAAGTTGACTCTGCTCAGAGACAGCTTGA-----AACCAAC	3145
Db	208	TCGGAGACGAGATCAAGAGCTTGAAAGAAATGACACAAGATGTTTCACAACAGACCAACTGGA	267
QY	3146	TCAACTCTCTGAACAATTCGGAACAACGTAAGCTCAGAACTCTCACTGTGTTCTTCAACAGA	3205
Db	268	GCAACACATCATCTACCTACCAAGAACGTGGAGCCACCACTATGGCTTCAACAAGAACCC	327
QY	3206	AGAAAGAACCGGAGATGCGCGAGACAGAGAACACTCTGGAGAACCAAGATTTCTTCA	3265
Db	328	TGACCCGAGGGCAACACCATTCACAGGAGCCATGGCCCAAGTTCAAGAGACAGTTCCTCG	387
QY	3266	AACACTCAAAAGGCGCTCGTCACATATTATATATGAGCGAGTCTCTCTGTGAAGACTCTCT	3325
Db	388	ACCGCGACATCAAGTTGTGACAGCTTACCTGGACACCCACTGACCGCCACGACAGGTGACCA	447
QY	3326	CCGAGAGAGCATTCAGACCGAGAGATTACTACGCCACGCTCGAACAATTCAAGTCTGT	3385
Db	448	GCAAGGAGCGCGTGATCTCGAAGGTCAACGCTCCCAAGCGGCAAGGGCAGACCAACCCCA	507
QY	3386	CTAAGCTCGAAGGCAAGCTGTAAGGCAACCTGAGACTGTGAGAAAGAAAGAACTCAGCTACC	3445
Db	508	CCAAAGCGCGCGTGAATCTCTGAACAACAGGAGATTACAAAGATGCTGATCGACAACGGCTACA	567
QY	3446	T 3446	
Db	568	T 568	

RESULT 54
US-08-471-033-45
: Sequence 45, Application US/08471033
: Patent No. 5770656
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Koziel, Michael G

```

? APPLICANT: Mullins, Martha A
? APPLICANT: Nye, Gordon J
? APPLICANT: Carr, Brian
? APPLICANT: Desai, Nalini M
? APPLICANT: Kostichka, N. Kristy
? APPLICANT: Duck, Nicholas B
? APPLICANT: Estruch, Juan J
? TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
? NUMBER OF SEQUENCES: 50
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CIBA-GEIGY Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: NY
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30B
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/471,033
? FILING DATE:
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/314,594
? FILING DATE: 09-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/218,018
? FILING DATE: 23-MAR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/037,057
? FILING DATE: 25-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Pace, Gary M.
? REGISTRATION NUMBER: P-40,403
? REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8582
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 45:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1358 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Synthetic DNA"
? HYPOTHEICAL: NO
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 9..1355
? OTHER INFORMATION: /note= "Maize optimized VIP2A(a)
? OTHER INFORMATION: with the Bacillus secretion signal removed and the vacuola
? OTHER INFORMATION: targeting signal inserted as contained in pcib5533"
US-08-471-033-45
Query Match 0.98; Score 43; DB 1; Length 1358;
Best Local Similarly 44.7%; Pred. No. 0.21;
Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;
```

QY	2972	AGAAAGCTGGAAAGGAGCATCAATTAACCTCAAGAGACATCTGCACCTCTCGACCTT	3033
Db	205	AGAAAGGAGGAAGTGGAAAGCTTACCGGCCACGAGAGAGGCGCAGATGAACCACTTCTCGG	264
QY	3032	ACAAACAAGTCCAAACTGCAACCTGAGAGACTCTTGCACAGAAGAACGACATGCGGCAAGT	3091
Db	265	ACAAACAAGACGACATCAAGACCAACTACAGAGAGATCCCTTCAGCATBAGCGCCACT	324
QY	3092	ATAAGATGCAGATCAAGAAAGTTGCACCTCTCAAGGAGCAGCTTGA-----AAGCAAC	3145
Db	325	TCGAGGACGAGATCAAGGACCTCAAGGAAGATGACCAAGATGTTCCACACAGAACCACTCGA	384


```

1  APPLICANT: Kostichka, N. Kristly
2  APPLICANT: Duck, Nicholas B
3  APPLICANT: Estruch, Juan J
4  TITLE OF INVENTION: No. 5849870e1 Pesticidal Proteins and Strains
5  NUMBER OF SEQUENCES: 50
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: CIBA-GEIGY Corporation
8  STREET: 7 Skyline Drive
9  CITY: Hawthorne
10 STATE: NY
11 COUNTRY: USA
12 ZIP: 10532
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.30B
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/463,483A
21 FILING DATE:
22 CLASSIFICATION: 530
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 08/314,594
25 FILING DATE: 09-SEP-1994
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 08/218,018
28 FILING DATE: 23-MAR-1994
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 08/037,057
31 FILING DATE: 25-MAR-1993
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Spittell, W. Murray
34 REGISTRATION NUMBER: 32,943
35 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: 919-541-8615
38 TELEFAX: 919-541-8689
39 INFORMATION FOR SEQ ID NO: 45:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 1358 base pairs
42 TYPE: nucleic acid
43 STRANDEDNESS: single
44 TOPOLOGY: linear
45 MOLECULE TYPE: other nucleic acid
46 DESCRIPTION: /desc = "synthetic DNA"
47 HYPOTHETICAL: NO
48 FEATURE:
49 NAME/KEY: CDS
50 LOCATION: 9...1355
51 OTHER INFORMATION:
52 OTHER INFORMATION: /note= "Maize optimized VIP2A(a)
53 with the Bacillus secretion signal removed and the vacuolar
54 targeting signal inserted as contained in pCIB5533"
55
56 US-08-463-483A-45

```

	Query Match	0.9%	Score 43	DB 2	Length 1358	
	Best Local Similarity	44.7%	Pred. No. 0.21			
	Matches	215	Conservative	0	Mismatches	260
					Indels	6
					Gaps	1
QY	2972	AGAACTCGAAGAGACATCATTAAGCTGAGAGAACACTGCACACTGAGCTTGCACCTGT	3031			
Db	205	AGAAAGGAAAGAGCTGAGAGCTTACCGCCACCGAAGAGGCAAGATGAAACAACCTTCTGG	264			
QY	3032	ACAACAATGACAACTGAACTGGGAGACACTCTTGACAGAAAGAAAGACAGTGGCGCAAGT	3091			
Db	265	ACAAACAGACGACATCTCAAGACCACTCAAGAGAGATACCTTCAGCATAGCGGCGACGT	324			
QY	3092	ATAAGATCAGATCAAGAAGTTGACTTGCCTCAAGAGACAGCTTGA-----AAGCAAAC	3145			
Db	325	TCGAGAGCAGAGATCAAGGACCTGAGAGGAGATGACAAAGATGTTGCAACAAGACCAACTGA	384			
QY	3146	TCAACTCACTGAACAATCCGAAACACGTAATCTGCAACTCTCTCACTGTTCTTCAACAAGA	3205			
Db	385	GCAACAGCATATATCTCTACAAAGACGTGGAGCCCAACCACTATGCGTTCAACAAGAGCC	444			

OY	3206	AGAAAGGAAGCGGATGTGCCGACACAGAGAACAACCTCTGGAGAACCAAGATTCTTTC	3265
Dd	445	TGACCAGAGGGCACACCATATAACAGCAGCCACTGGCCAGTTCCAAGACAGCTTGCG	504
OY	3266	AACACTACAAGAAGCGCTGCACGATTATTATTAATGGCAGATCTTCTCCTGGAAGACTCT	3325
Dd	505	ACCCGCAATCAAGTTGACAGCTACTCTGGACACCCACTGACCCGCCAGCGGTGACCA	564
OY	3326	CCGAGAGGAGCATTCACAGCCGAGGATTAAGTACGCCAGCTCGAGAACTTCGAAGTCTGT	3385
Dd	565	GCAAGGACCGGTGATCTCTAAGGTGACCGTCCCCAGCGGCAAGGGCAGCACACCCCA	624
OY	3386	CTAAGCTCGAAGGCAAGCTGAAGCAACCTGAACCTGGAGAAGAAGTCAAGCTAC	3445
Dd	625	CCAAGCGCCGGGTGATCTCTGAACAACAGCAGATACAAAGTCTGATCGACACGCGCTACA	684
OY	3446	T 3446	
Dd	685	T 685	

RESULT 57
 US-08-471-046A-45
 Sequence 45, Application US/08471046A.
 Patent No. 5866326
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozien, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desail, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
 TITLE OF INVENTION: Protein Genes
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5866326artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,046A
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLW4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 45:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1358 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "synthetic DNA"
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..1355 /note= "Maize optimized VIP2A(a)"
 OTHER INFORMATION: with the Bacillus secretion signal removed and the vacuolar
 OTHER INFORMATION: targeting signal inserted as contained in pcIB5533"
 US-08-471-046A-45

Query Match 0.98; Score 43; DB 2; Length 1358;
 Best Local Similarity 44.7%; Pred. No. 0.21;
 Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 2972 AGAAGCTGGAAGAGGAGCATATTAAGCTGGAAGAGACACTGCAAGTGGCTTGACCTGT 3031
 DB 205 AGAAGGAGAGAGGAGTGAAGCTTACCGCCACCGAGAGGGCAAGATGAAACACTTCTCTG 264
 QY 3032 ACAACAAGTACAACTGAACTGGAGAGACTCTTCGACAAAGAGAGACAGTGGCAAGT 3091
 DB 265 ACAACAAGAGAGCATCAAGACCACTACAAAGAGATCACCTTCAGATATAGCGGGAGCT 324
 QY 3092 ATAAAGATGCAGATCAAGAACTTGAATCTGCTCAAGAGGAGAGCTTGA-----AAGCAAC 3145
 DB 325 TCGAGGACGAGATCAAGAGACTGAAGAGATGCAAGAGATGTCGACCAAGACCAACCTGA 384
 QY 3146 TCAACTCAGTGAACAACTCCGAACAGTACTGAGAACTTCTCAGTGTCTTCAACAGA 3205
 DB 385 GCAACAGCATCATCACCCTACAAAGAGCTGAGGCCACCAACATTCGGTTCAACAAGAGCC 444
 QY 3206 AGAAGGAGAGCGGAGATGCGCGGAGAGACAGAGAACTCTGGAGAAACCAAGATTCTTCTCA 3265
 DB 445 TGACCGAGGGCAACACCATCAACAGCGCCATGGCGCCAGTTCAAGAGACAGTCTCTCTG 504
 QY 3266 AACACTACAAAGGCTCTGTAATTTATATGCGGAGTCTTCTCCTCTGAACACTCTCT 3325
 DB 505 ACCGACACATCAAGATTCGAGAGCTTACCTGGACACCCACCTGACCGCCACAGAGTGAGCA 564
 QY 3326 CCGAGGAGAGCATCGAAGCAGAGATTAACGCGAGATTAACGCGACCTTCGAACTTCAAGTCTCTGT 3385
 DB 565 GCAAGGAGCGCGTGAATCTTGAAGGTGACCTCCCGAGCGCAAGGGCAGACCAACCTCCA 624
 QY 3386 CTAAAGCTGGAAGGCAAGCTGGAAGAGCAACCTGGAACCTGGAAGAGAAAGAACTGACCTACC 3445
 DB 625 CCAAGGCGGGCGTGAATCTTGAACAGAGCAAGTATCAAGAGATGCTGATGAGCAACGGCTACA 684
 QY 3446 T 3446
 DB 685 T 685

RESULT 58
 US-08-470-566B-45
 Sequence 45, Application US/08470566B
 Patent No. 5872212
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Koziele, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No. 5872212e1 Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5872212artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/470,566B
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Weigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 45:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1358 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: other nucleic acid
 DESCRIPTION: /desc = "synthetic DNA"
 HYPOTHETICAL: NO
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 9..1355 /note= "Maize optimized VIP2A(a)"
 OTHER INFORMATION: with the Bacillus secretion signal removed and the vacuolar
 OTHER INFORMATION: targeting signal inserted as contained in pcIB5533"
 US-08-470-566B-45

Query Match 0.98; Score 43; DB 2; Length 1358;
 Best Local Similarity 44.7%; Pred. No. 0.21;
 Matches 215; Conservative 0; Mismatches 260; Indels 6; Gaps 1;

QY 2972 AGAAGCTGGAAGAGGAGCATATTAAGCTGGAAGAGACACTGCAAGTGGCTTGACCTGT 3031
 DB 205 AGAAGGAGAGAGGAGTGAAGCTTACCGCCACCGAGAGGGCAAGATGAAACACTTCTCTG 264
 QY 3032 ACAACAAGTACAACTGAACTGGAGAGACTCTTCGACAAAGAGAGAGACAGTGGCAAGT 3091
 DB 265 ACAACAAGAGAGCATCAAGACCACTACAAAGAGATCACCTTCAGATATAGCGGGAGCT 324
 QY 3092 ATAAAGATGCAGATCAAGAACTTGAATCTGCTCAAGAGAGAGCTTGA-----AAGCAAC 3145
 DB 325 TCGAGGACGAGATCAAGAGACTGAAGAGATGCAAGAGATGTCGACCAAGACCAACCTGA 384
 QY 3146 TCAACTCAGTGAACAACTCCGAACAGTACTGGAAGAACTTCTCAGTGTCTTCAACAGA 3205
 DB 385 GCAACAGCATCATCACCCTACAAAGAGCTGAGAGCCACCAACATTCGGCTTCAACAAGAGCC 444

[illegible]

```

RESULT 59
US-08-469-334-45
: Sequence 45: Application US/08469334
: Patent No. 5990383
: GENERAL INFORMATION:
: APPLICANT: Warren, Gregory W
: APPLICANT: Kozziel, Michael G
: APPLICANT: Mullins, Martha A
: APPLICANT: Nye, Gordon J
: APPLICANT: Carr, Brian
: APPLICANT: Desai, Nalini M
: APPLICANT: Kostichka, N. Kristy
: APPLICANT: Duck, Nicholas B
: APPLICANT: Estruch, Juan J
: TITLE OF INVENTION: No. 5990383el Pesticidal Proteins and Strains
: NUMBER OF SEQUENCES: 50
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30B
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,334
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/463,483
: FILING DATE:
: APPLICATION NUMBER: US 08/314,594
: FILING DATE: 09-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/218,018
: FILING DATE: 23-MAR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/037,057
: FILING DATE: 25-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Spullin, W. Murray
: REGISTRATION NUMBER: 32,943
: REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8615
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 45:

```

```

?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1358 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: other nucleic acid
?      DESCRIPTION: /desc = "Synthetic DNA"
?      HYPOTHETICAL: NO
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 9..1355
?      OTHER INFORMATION:
?      OTHER INFORMATION:
?      OTHER INFORMATION:
US-08-469-334-45      /note= "Maize optimized VIP2A(a)
                        with the Bacillus secretion signal removed and the vacuolization
                        targeting signal inserted as contained in pCIB533"

```

Query Match	0.9%	Score 43	DB 2	Length 1358
Best Local Similarity	44.7%	Pred. No	0.21	
Matches 215	Conservative	0	Mismatches 260	Indels 6
				Gaps 1

QY	2972	AGAAAGCTGGAAGAGGACATTCATTAATTAACCTGTAAGAAGACATCTGCATAGCTGACCTG	3033
Db	205	AGAAAGGGAAGAGAGTGGAAAGCTTACCGCCACCGAGAGAAGGGCAAGATGAAACAACTTCTCG	264
QY	3032	ACAACAAGTACAAACTGGAACCTGGAGAGACTCTTCGACAAAGAAGACACTGGGCAAGT	3091
Db	265	ACAACAAGAACGACATCAAGACCACACTACAGAGATCACCTTCAGCATATAGCCGGCAGACT	324
QY	3092	ATAAGATGACGATTCACGAAGAGTTGACTCTCTCAAGAGCAGCTTGA-----AACCAAC	3145
Db	325	TCGAGGACGAGATTCAAAGACCTGGAAGAAATGACAAAGTGTTCACCAAGACCAACTGGA	384
QY	3146	TCAACTCTAGTAAACAATCCGAAACACGTAAGTACGAAACTTCTCAGTGTCTTCAACAAGA	3205
Db	385	GCAACACGATCATCACTACCTACAAAGAACGTGGAGCCCAACCCATCCGCTTCAACAAGACCC	444
QY	3206	AGAAAGGAAGCGGATTCGGCCGAGACGAGAAACACTCTGGAGAACAACAAGATTCTTCTCA	3265
Db	445	TGACCGAGGGCAACACCATTAACACAGGACGCCATGGCCAGTTCAAGAGACAGTTCCTGG	504
QY	3266	AACACTACAAAGGCGCTCGTCAAGTATTATTAATGGCGAGTCTTCTCTCTGGAAGACTCTCT	3325
Db	505	ACGCGCGACATCAAGTTTCAGACAGTACTACCTGGACACCCACTGACCCGCCAGAGGGTGAACA	564
QY	3326	CCGAGGAGACATTCAGACCGGAGATTAACCTAGCCAGCTCGAGAACAATTCAAGGTCCTGT	3385
Db	565	GCAAGGAGCGCGTGAATCTCTGAAGGTACCGTCCCAACGCGCAAGGGCAGCACACCCCA	624
QY	3386	CTAAGCTCAGAGGCAAGCTGAAGAGCAACACTGAACCTGGAGAAAGAAGACTCAGCTACC	3445
Db	625	CCAAGGCGCGGCTGATCTCTGAACAACACAGGAGTACAAAGATCTGTATCGACAACGGCTACA	684
QY	3446	T 3446	
Db	685	T 685	

Search completed: March 31, 2003, 04:33:03
Job time : 559 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 23:49:26 ; Search time 4855 Seconds

(without alignments)
16479.033 Million cell updates/sec

Title: US-09-269-874A-2

Perfect score: 4940

Sequence: 1 cgcacgcgtatgaataatcat.....ttcatctaatagatcgtatg 4940

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 180 summaries

Database :

EST:*
1: em_estbda:*
2: em_estbda:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	235	4.8	537	13	B1815756 PFESTRoaa3
2	220.8	4.5	600	14	BQ452315 PFESTRoaa9
3	194	3.9	500	9	A0088128
4	192.4	3.9	491	14	BQ597179 PFESTRoab3
5	179.8	3.6	432	14	BQ633144 PFESTRoab3
6	168.8	3.4	500	9	A0086832

C	7	161.2	3.3	501	13	B1814631	B1814631 PFESTRoaa4
	8	160.4	3.2	422	14	BQ451709	BQ451709 PFESTRoab0
	9	160	3.2	522	14	BQ452445	BQ452445 PFESTRoaa9
	10	159	3.2	455	14	BQ451036	BQ451036 PFESTRoab0
	11	158	3.2	500	9	A0087573	A0087573 PFESTRoab0
	12	156.4	3.2	500	9	A0088129	A0088129 PFESTRoab0
	13	134.4	2.7	303	13	B1815239	B1815239 PFESTRoaa1
	14	132.4	2.7	419	14	BQ451367	BQ451367 PFESTRoaa9
	15	124.4	2.5	408	13	B1814931	B1814931 PFESTRoaa0
	16	124.4	2.5	418	14	BQ451572	BQ451572 PFESTRoab0
	17	119.8	2.4	365	14	BQ451661	BQ451661 PFESTRoab0
	18	112.2	2.3	597	13	B1815957	B1815957 PFESTRoaa1
	19	108.8	2.2	500	9	A0086246	A0086246 PFESTRoaa1
	20	108.8	2.2	755	13	B1815731	B1815731 PFESTRoaa1
	21	108.8	2.2	753	13	B1816043	B1816043 PFESTRoaa1
	22	107.6	2.2	678	13	B18165725	B18165725 PFESTRoaa1
	23	105	2.1	402	13	B181670649	B181670649 PFESTRoaa1
	24	102.6	2.1	400	14	N97689	N97689 PFESTRoaa1
	25	101.6	2.1	354	14	N97605	N97605 PFESTRoaa1
	26	98.2	2.0	647	17	A2526918	A2526918 PFESTRoaa1
	27	97.2	2.0	660	17	B1815472	B1815472 PFESTRoaa1
	28	91.2	1.8	660	17	B18170928	B18170928 PFESTRoaa1
	29	89.6	1.8	732	13	B18160106	B18160106 PFESTRoaa1
	30	89.6	1.8	313	14	T18122	T18122 PFESTRoaa1
	31	88.4	1.8	362	14	BQ451205	BQ451205 PFESTRoaa1
	32	86.2	1.7	352	14	BQ573302	BQ573302 PFESTRoaa1
	33	85.2	1.7	282	14	T18003	T18003 PFESTRoaa1
	34	82.8	1.7	79	13	B18153470	B18153470 PFESTRoaa1
	35	80.6	1.6	878	17	B18165179	B18165179 PFESTRoaa1
	36	79	1.6	483	13	B18167609	B18167609 PFESTRoaa1
	37	79	1.6	769	13	B18166688	B18166688 PFESTRoaa1
	38	79	1.6	503	13	B18159366	B18159366 PFESTRoaa1
	39	78.8	1.6	344	13	B2413205	B2413205 PFESTRoaa1
	40	77.2	1.6	540	13	B2413205	B2413205 PFESTRoaa1
	41	74.4	1.5	948	13	B2415494	B2415494 PFESTRoaa1
	42	73.8	1.5	1885	17	BE420745	BE420745 PFESTRoaa1
	43	73.4	1.5	1885	17	BE420745	BE420745 PFESTRoaa1
	44	71.6	1.4	691	17	BH040136	BH040136 PFESTRoaa1
	45	69.8	1.4	682	13	B18170771	B18170771 PFESTRoaa1
	46	68.6	1.4	908	17	A2548467	A2548467 PFESTRoaa1
	47	68.6	1.4	544	13	B2420171	B2420171 PFESTRoaa1
	48	68.4	1.4	596	17	A2460541	A2460541 PFESTRoaa1
	49	68.4	1.4	907	17	CNS02134	CNS02134 PFESTRoaa1
	50	68	1.4	494	13	B18160881	B18160881 PFESTRoaa1
	51	67.2	1.4	500	9	A0087665	A0087665 PFESTRoaa1
	52	67.2	1.4	827	17	CNS02156	CNS02156 PFESTRoaa1
	53	66.4	1.3	1168	14	BQ931326	BQ931326 PFESTRoaa1
	54	66	1.3	731	17	A2191902	A2191902 PFESTRoaa1
	55	65.8	1.3	960	17	CNS005F3	CNS005F3 PFESTRoaa1
	56	65.6	1.3	626	13	B2426817	B2426817 PFESTRoaa1
	57	65.4	1.3	783	17	CNS01112	CNS01112 PFESTRoaa1
	58	65.4	1.3	897	17	CNS07AB2	CNS07AB2 PFESTRoaa1
	59	65.4	1.3	1135	17	CNS033C0	CNS033C0 PFESTRoaa1
	60	65.4	1.3	1055	14	BQ876453	BQ876453 PFESTRoaa1
	61	65.2	1.3	885	17	A2541857	A2541857 PFESTRoaa1
	62	65	1.3	641	17	AQ946120	AQ946120 PFESTRoaa1
	63	64.8	1.3	1101	17	CNS00K85	CNS00K85 PFESTRoaa1
	64	64.8	1.3	500	9	A0086975	A0086975 PFESTRoaa1
	65	64.4	1.3	843	17	A2551618	A2551618 PFESTRoaa1
	66	64.4	1.3	1223	17	B12981	B12981 PFESTRoaa1
	67	64.2	1.3	517	13	B18163917	B18163917 PFESTRoaa1
	68	64	1.3	1309	13	B2463041	B2463041 PFESTRoaa1
	69	64	1.3	976	17	AG136173	AG136173 PFESTRoaa1
	70	63.8	1.3	650	10	AV995927	AV995927 PFESTRoaa1
	71	63.6	1.3	1084	17	AG032755	AG032755 PFESTRoaa1
	72	63.6	1.3	517	13	B18163917	B18163917 PFESTRoaa1
	73	63.4	1.3	885	17	A2542280	A2542280 PFESTRoaa1
	74	63.2	1.3	356	13	BQ337970	BQ337970 PFESTRoaa1
	75	63	1.3	910	13	B2415636	B2415636 PFESTRoaa1
	76	63	1.3	932	17	A2689160	A2689160 PFESTRoaa1
	77	62.8	1.3	443	6	BW374465	BW374465 PFESTRoaa1
	78	62.8	1.3	1300	13	B2466018	B2466018 PFESTRoaa1
	79	62.8	1.3	1300	13	B2466018	B2466018 PFESTRoaa1

80	62.6	1.3	603	13	BJ366946
81	62.6	1.3	997	17	CNS0134P
82	62.6	1.3	1036	17	CNS031MUT
83	62.4	1.3	394	9	AU060224
84	62.4	1.3	439	13	BM160668
85	62.4	1.3	527	13	BM160302
86	62.4	1.3	530	13	BM166839
87	62.4	1.3	784	13	BM166681
88	62.2	1.3	1101	17	CNS0039G
89	62.2	1.3	804	17	AG077527
90	61.8	1.3	848	17	AZ672705
91	61.8	1.3	870	17	AQ330286
92	61.6	1.2	671	17	AZ749598
93	61.6	1.2	681	17	CNS02EOD
94	61.4	1.2	538	13	BJ417399
95	61.2	1.2	617	13	BJ416228
96	61.2	1.2	839	17	CNS0122R
97	61	1.2	359	13	BJ366856
98	61	1.2	867	17	CNS0054A
99	61	1.2	871	17	AZ671726
100	60.8	1.2	576	9	AU060772
101	60.8	1.2	634	17	AZ008814
102	60.8	1.2	783	17	CNS011P
103	60.8	1.2	1101	17	CNS0153P
104	60.6	1.2	744	17	AG044338
105	60.6	1.2	877	17	AZ531291
106	60.6	1.2	912	17	AZ551092
107	60.4	1.2	648	10	AV676742
108	60.4	1.2	913	17	BM158934
109	60.4	1.2	997	17	CNS005TE
110	60.2	1.2	461	17	AG072942
111	60	1.2	950	13	BJ371297
112	60	1.2	834	17	BI2387
113	60	1.2	964	13	BM416130
114	60	1.2	966	13	BM415686
115	60	1.2	1184	17	BI3117
116	60	1.2	1398	13	BM416038
117	59.8	1.2	627	10	BJ426316
118	59.8	1.2	688	10	AV671840
119	59.8	1.2	884	17	AZ199419
120	59.8	1.2	884	17	AG056518
121	59.8	1.2	896	13	BM415641
122	59.8	1.2	924	13	BM415947
123	59.6	1.2	1008	17	AG134526
124	59.4	1.2	538	17	AZ064820
125	59.4	1.2	582	13	BJ369186
126	59.4	1.2	697	17	AZ374916
127	59.4	1.2	787	17	CNS007XT
128	59.4	1.2	959	17	CNS00655
129	59.4	1.2	1491	12	BE882936
130	59.2	1.2	877	17	AZ548629
131	59	1.2	858	17	CNS0127J
132	59	1.2	1185	12	BF273407
133	58.8	1.2	814	17	AZ672545
134	58.8	1.2	947	17	AG056380
135	58.8	1.2	954	14	BO670925
136	58.6	1.2	613	13	BJ360997
137	58.6	1.2	640	13	AZ476780
138	58.6	1.2	784	9	AL5525973
139	58.6	1.2	931	17	AG446286
140	58.6	1.2	952	17	CNS014BF
141	58.4	1.2	803	17	BI1974
142	58.4	1.2	865	17	AQ324474
143	58.4	1.2	1007	17	CNS006X5
144	58.2	1.2	421	17	AZ113646
145	58.2	1.2	850	17	AZ186328
146	58	1.2	546	14	CZ2974
147	58	1.2	784	14	BO962626
148	58	1.2	861	17	AZ137643
149	58	1.2	866	17	AG073614
150	58	1.2	934	13	BM415088
151	58	1.2	979	13	BM415502
152	58	1.2	1204	13	BM452144

153	57.8	1.2	539	9	AU076365
154	57.8	1.2	542	13	BM167382
155	57.8	1.2	816	17	CNS02KJV
156	57.8	1.2	878	17	CNS0008X
157	57.8	1.2	886	17	CNS075DF
158	57.8	1.2	931	17	BM160272
159	57.8	1.2	952	17	AO897460
160	57.8	1.2	1042	17	CNS0148K
161	57.8	1.2	1059	17	CNS0022X
162	57.8	1.2	1079	17	CNS0351X
163	57.8	1.2	1101	17	CNS001JT
164	57.6	1.2	297	13	BM633446
165	57.6	1.2	608	13	BJ330059
166	57.6	1.2	617	17	AZ296172
167	57.6	1.2	735	17	AZ185710
168	57.6	1.2	888	17	CNS027EX
169	57.6	1.2	964	17	AZ667479
170	57.6	1.2	1215	14	BM922560
171	57.4	1.2	739	14	BO991717
172	57.4	1.2	829	17	CNS011NU
173	57.2	1.2	571	17	AZ952930
174	57.2	1.2	792	17	AG032727
175	57.2	1.2	796	17	CNS0118D
176	57.2	1.2	944	17	CNS00A5G
177	57.2	1.2	1139	17	AO897537
178	57.2	1.2	1382	14	BM609864
179	57	1.2	645	9	AI389106
180	57	1.2	870	17	AO866797

ALIGNMENTS

AU076365 AU076365
 BM167382 EST569905
 AL201604 Tetradon
 AL097371 Drosophila
 AL429961 Clone BAO
 BH160272 ENT0497R
 AO897460 HS_3134.A
 AL103838 Drosophila
 AL097133 Drosophila
 AL28786 Tetradon
 AL068307 Tetradon
 BM633446 170006875
 BJ330059 BJ330059
 AZ296172 RPCT-23-1
 AZ185710 SP_1005.B
 AL184578 Tetradon
 AZ667479 ENTJY62TF
 BM922560 AGENCOURT
 BO991717 QGF23J24.
 AL100500 Drosophila
 AZ952930 2M0217N19
 AG032727 Pan trogl
 AL099943 Drosophila
 AL054949 Drosophila
 AO897537 HS_3153.A
 BM609864 AGENCOURT
 AI389106 GH20192.5
 AO866797 nbe00029E

RESULT 1
 BI815756
 LOCUS
 DEFINITION B1815756 537 bp mRNA linear EST 03-OCT-2001
 PFEST0a32c09.y1 Plasmodium falciparum 3D7 asexual cDNA plasmodium
 falciparum cDNA 5' similar to TR:Q9TV68 Q9TV68 MSAL PROTEIN ;, mRNA
 sequence.

ACCESSION
 BI815756
 VERSION
 B1815756.1 GI:15909334

KEYWORDS
 malaria parasite P. falciparum.

SOURCE
 ORGANISM
 Plasmodium falciparum

REFERENCE
 1 (bases 1 to 537)
 Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
 Maitra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,
 Tsagaris, V., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
 Waterston, R., Wilson, R. and Sibley, D.

AUTHORS
 Unpublished (2001)

TITLE
 JOURNAL
 Contact: L. David Sibley

COMMENT
 Washt Plasmodium EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

EMAIL: estewatson.wustl.edu
 Library was constructed by Deoban Chakrabarti DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: L. David Sibley
 (sibley@borcim.wustl.edu), Washington University
 Seq primer: -40UP from Gibco
 High quality sequence stop: 422.

Location/Qualifiers
 1..537
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
 /lab_host="DH10B (genehog, Invitrogen, Inc.)"

FEATURES

source

/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 Zapit vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

BASE COUNT 217 a 75 c 83 g 162 t
 ORIGIN

Query Match 4.8%; Score 235; DB 13; Length 537;
 Best Local Similarity 66.0%; Pred. No. 9.3e-43;
 Matches 356; Conservative 0; Mismatches 180; Indels 3; Gaps 1;

QY 4401 AACATCCAGACAGCTGCGAGATTTCAGAAAAATACATTTCTCGAATTGCAGA 4460
 Db 1 AGCAATTGATGACAAATAGATCTTTTAAACCCTTACGACTTCGAGCAATTTAAAA 60
 QY 4461 CCTGCTACCATTTTAACACACATCTCTGACCAAGTTTCTCCACTGCGATGCT 4520
 Db 61 ATTGTAATATATATACGAAAAAAGATATGCTGGCAAAATTACTTACAGGATAGT 120
 QY 4521 GTTCGAAAACCTGGCAAAACAGTCTGAGCAATCTGCGACGCGACCTGCGAGGAT 4580
 Db 121 TC-----AAATTTTCTTAATACATTAATATCAAAATTTAATGAGAAATTCAGATAT 177
 QY 4581 GCTGAACATCTCCAGACCAATGGGTGAGAAAGACGTCCTCCCAATAGCGGCTGTTT 4640
 Db 178 GTTAAACATTTTACACACCAATGGGTGAGAAAGACGTCCTCCCAATAGCGGCTGTTT 237
 QY 4641 CAGCATCTGACAGCGCCGAGAGAGTGTCTCTGAACTTCAACAAAGAAAGAGA 4700
 Db 238 CAGCATTTTAATGAGAAAGAGATGTAAATGTTTAAATTAATTAACAAAGAGGTGA 297
 QY 4701 TAAATGCTGGAGAACCCCAACCTACTGCAATGAGAAACATAGCGGGGTGAGCGCGA 4760
 Db 298 TAAATGCTTGAATAATCCAAATCTACTTGTACGAAAAATATGATGATGATGCGAGA 357
 QY 4761 TGTAAATGACGAGAGAGAGAGCGGCTCTAACGAGAAAGAAATCAGATGCGAGTAC 4820
 Db 358 TGCCACATGTACCGAAGAAAGATTCAGTAGCAGCAAGAAAGAAATCAGATGATGATAC 417
 QY 4821 TAAACCCGACCTCTATCCACCTCTGACAGGATTTTTCGTCACGCTTAATTTCTCTGG 4880
 Db 418 TAAACCTGATTTCTATTCACCTTTTCGATGATGATTTTTCGACGCTTCTCTCTTGA 477
 QY 4881 CATCTCTCTGCTGATGCTCATGCTCTGATGATGCTTCAATCTAATGATGATGATG 4939
 Db 478 AATATCATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 536

RESULT 2
 BQ452315
 LOCUS 600 bp mRNA Linear EST 29-MAY-2002
 DEFINITION Pfesr0a95a12.y1 plasmodium falciiparum 3D7 asexual cDNA Plasmodium
 SURFACE PROTEIN 1 PRECURSOR ; mRNA sequence.
 ACCESSION BQ452315
 VERSION BQ452315.1 GI:21255427
 KEYWORDS EST.
 SOURCE malaria parasite P. falciiparum.
 ORGANISM Plasmodium falciiparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 REFERENCE 1 (bases 1 to 600)

AUTHORS Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
 Maria, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Thelshing, B.,
 Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,
 Tsagarelis, V., R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,
 Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
 Waterston, R., Wilson, R., and Sibley, D.
 TITLE Washu Plasmodium Est Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: L. David Sibley
 Washu Plasmodium Est Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 Library was constructed by Debopam Chakrabarti DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: L. David Sibley
 (sibley@borcim.wustl.edu), Washington University
 Seg primer: -400P from Gibco
 High quality sequence stop: 432.
 Location/Qualifiers
 1..600
 /organism="Plasmodium falciiparum"
 /db_xref="taxon:5833"
 /clone.lib="Plasmodium falciiparum 3D7 asexual cDNA"
 /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
 /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 Zapit vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

FEATURES

Source

BASE COUNT 276 a 79 c 72 g 173 t
 ORIGIN

Query Match 4.5%; Score 220.8; DB 14; Length 600;
 Best Local Similarity 63.6%; Pred. No. 1.7e-39;
 Matches 336; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 3064 TTGCAAGAGAGAGACAGTGGCGAAGTATAGACAGTCAAGATGATGATGCTGCTC 3123
 Db 1 TTTAATAGAAAAAGAACTGGCCAGACAAAGCAAAATTAATAAACTTATATTA 60
 QY 3124 AAGAGACGCTTGAAGAGAACTCACTGACGAACTCGAAACAGCTACTGAGAAC 3183
 Db 61 AAAGACATTAAGATCAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 120
 QY 3184 TTCTCAGTCTTCTTCAACAGAGAGAGCGGATGCGGAGACAGACACCTCTG 3243
 Db 121 TTTTCTGTTTCTTAAACAAAAAGAGCTGAAATAGCAGAACTGAAACACATTA 180
 QY 3244 GAGAACACCAAGATTTCTTCAACACATCAAGAGCGCTGCTCAATTAATTAATGCGAG 3303
 Db 181 GAAACACAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240
 QY 3304 TCTTCTCCTGAGAGACTCTCTCCAGAGAGACATCCAGACGAGATTAACGCCAGC 3363
 Db 241 TCATCTCCATTAATAAATCTTGAAGATTAATTAATTAATTAATTAATTAATTAATTAAT 300
 QY 3364 CTCGAGAACTTCAAGGCTCTCTAAGCTCGAAGGCAAGTGAAGACACCTGAGACCTG 3423
 Db 301 TTAGAAAAATTTAGAGTATTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTA 360

OY 3424 GAGGAAGAAGAACTGACGTACCCTCCTGACGGACATCACTCATGCGCCAGTCAAG 3483
| | | | | | | | | | | | | | | | | | | |
Db 361 GGAAAAAATAATTATCTTTCTTATTCACMGATTTACATCATTTATTTACTGAATTAATA 420
| | | | | | | | | | | | | | | | | | | |
OY 3484 GAAGTCATTAAACAAGAACAATCACCGCAATPACCCAGAGAAATPATACAGACGTG 3543
| | | | | | | | | | | | | | | | | | | |
Db 421 GAAGTATATATTAATAATTAATAATTTATACAGTAATTTCTCCAAGTAAAATATATAGAAAGTT 480
| | | | | | | | | | | | | | | | | | | |
OY 3544 AATAACGCACCTGGAATCTTACAAGAAAGTTCCGCCGTGAAGAACAAGAT 3591
| | | | | | | | | | | | | | | | | | | |
Db 481 AACGAAGCTTTAAATCTTACGAATAATTTCTCCCAAGAACAAAGATT 528

RESULT 3
AM088128

LOCUS	500 bp	mrna	linear	EST 27-JAN-2001
AU0088128				
AU0088128	Sugano	Malaria	cDNA library	<i>Plasmodium falciparum</i> 307 CDNA

for precursor of major merozoite surface antigens, mRNA sequence.

KEYWORDS EST. Plasmodium falciparum 3D7. SOURCE

Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
1 (bases 1 to 500)

TITLE	UULL-malaria: a database for a full-length enriched cdna library from human malaria parasite, Plasmodium falciparum
NUMERIC	Nucleic Acids Res 28(71): 70-71 (2001)

MEDLINE	20574754
COMMENT	Contact: Junichi Watanabe Institute of Medical Sciences

The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel.: 03-5540-5370

Fax: 81-3-5449-5410
Email: jwatanabe@nagae.ims.u-tokyo.ac.jp

5. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

```
source 1. .500
/organism="Plasmodium falciparum 3D7"
/dataset="0007"
```

```

/db_xref="taxon:36329"
/clone="XPFn6559"

```

```
BASE COUNT      229 a      /dev_stage="erythrocytic stage"
                69 c      56 g      146 t
```

Query Match 3.98; Score 194; DB 9; Length 500;

Matches	290;	Conservative	0;	Mismatches	160;	Indels	0;	Gaps	0;
---------	------	--------------	----	------------	------	--------	----	------	----

Db 5 AATGTGATTCACCTTAATACCACATATGTTACAAACCTTTCGTCTTTTAC 64

Qy 3202 AMGAGAAAGAACCGAGATCGCCGAGACAGAGAAACACTCTGGAGAAACCAAGATTCTT 3261

3262 CTCGAACGCTACAAAGGCGCTCTGTCAGTATTATTAATGGCGAGCTTCTCCTCGAAGACT 3321

Db 125 TTGAACATTATTAAGACTTGTTAAATATTATATGCGTAATCATCTCCATTAAAACT 184

Db 185 TTAAGTAGATCAATTCACAGACAGATTAATATGCGAATTTAGAGAAATTTAGAGTA 244

Qy	3382	CTGTTTAAAGCTCGAAGGCAAGCTGAAGGACAACTCGAACCTGGAGAGAAAGAAAGCTCAGC	3441
Db	245	TTTAACTAAAATATAGATGGAAGAACTCAATCATATATTTTACATTTTAGGAAAGAAAAAATTAATCT	304
Qy	3442	TACCTCTCTAGCGGACGATCATCTGATTCGCGGAGCTCAAGGAAGTCATTTAGAACAG	3501
Db	305	TTCTTATCAAGTGGATTTACATCATTTTATTTACTGATTTAAAAAGATTAATAAAAAATTA	364
Qy	3502	AACCTACCCGGCAATAGGCCAAGCGAGAAATATACAGACGTGAATTAACGCACTGGAATCT	3561
Db	365	AATTATACAGTAATTTCTCCAAAGTAAAAATATAAAGAAAGTTAACGAAGCTTTTAAAAATCT	424
Qy	3562	TACAGAAGATTCTGCTCGTGAAGGAACAGAT	3591
Db	425	TACGAAATTTTCTCTCCAGAGCAAAAGTT	454

RESULT 4

000770

LOCUS	BQ597179	491 bp	mRNA	linear	EST 24-JUN-2002
DEFINITION	F5E8Toab35h07.y1 Plasmodium falcipticum 3D7 asexual cDNA Plasmodium				

SURFACE PROTEIN 1 PRECURSOR ;, mRNA sequence.
 B0597179
 ACCESSION
 B0597179

KEYWORDS EST.
SOURCE malaria parasite *P. falciparum*.

REFERENCE
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 491)
Massé V, Côté P, Phelipponart D, Hammond D, Clifton C, Pano P,
Armstrong

Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Runko, I., MacFarland, D., Polunov, I., Franklin, C., Crow, A.

Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
Waterston, R., Wilson, R. and Sibley, D.

JOURNAL
Unpublished (2001)
COMMENT
Contact: L. David Sibley
Infect Dis Med Hyg 1999; 123: 1001-1002

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel.: 314 386 1000

Fax: 314 286 1810
Email: est@watson.wustl.edu
Ribeaux was associated by Deborah Chakrabarti, DNM recognized by

Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley

Seq primer: -40UP from Gibco
High quality sequence stop: 420.

```
source      1. .491 |
            /organism="Plasmodium falciparum"
```

```

/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"

```

XhoI; library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage

erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A⁺ RNA was isolated by the polyAT-Tract method.

Magnesium particles, directional cDNA libraries were constructed by oligo d(T) priming of poly(A)⁺ RNA (5mg

these DNA and minor levels of a major restriction enzyme and Deep
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.

ions were mass analysed using the unmodified helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA

was electroporated into DH10B cells."

JOURNAL MEDLINE
COMMENT
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
20574754
Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@med.s.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo Nakagawa, K., Maruyama, K., and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1. 500
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/clone="XFPn3463"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
Location/Qualifiers

BASE COUNT 229 a 51 c 52 g 146 t 22 others
ORIGIN

Query Match 3.4%; Score 168.8; DB 9; Length 500;
Best Local Similarity 60.2%; Pred. No. 1.2e-27;
Matches 266; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 2939 ATATCACTCTCTTAACGATGATCAAGTAAAGTAAAGCGGAGAGCATCATATAGC 2998
1 ANGNANATTCAGGTTTGACCGACACAAAAATGCAATTAAATGATTAATAAAT 60
QY 2999 TGAAGAAGACACTGACACTGAGCTGACCTGACACAAAGTAAAGTGAAGTGA 3058
61 TAAAGATPACTTACAGTTATCATTTGATTTAATTAATTAATTAATTAATTAAT 120
QY 3059 GACTCTTCACAGAAGAAAGACAGCTGCGCAAGTATGATGACATGACAGTGC 3118
11 TTTTATTTATTAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 180
Db 121 CATTTATTTATTAATAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 180
QY 3119 TCGTCAGAGAGAGAGCTGGAAGCAAGTCAAGTCAAGTCAAGTCAAGTCA 3178
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 240
Db 181 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 240
QY 3179 AGAAGCTTCAGTGTCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3238
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 300
Db 241 AAAAGCTTCAGTGTCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 3239 CTCTGAGAGACACAGAGATTTCTTCAACAGAGAGAGAGAGAGAGAGAG 3298
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 360
Db 301 CATTAAGAAACACAAATAATTAATTAATTAATTAATTAATTAATTAATTA 360
QY 3299 GGAGAGCTCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3358
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 420
Db 361 GTGATCATCTCCATTAAAGCTTTAAGTGAATATCAATCAACAGAGAGATTA 420
QY 3359 CCAGCCTCGAGAGCTCAAGGT 3380
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 442
Db 421 CCAATTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 442

RESULT 7

BI814631/c

LOCUS BI814631 501 bp mRNA linear EST 03-OCT-2001
DEFINITION Pf18toaa40c07.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
falciparum cDNA 5' similar to TR:09UA18 09UA18 SURFACE PROTEIN-1 ;,
mRNA sequence.

ACCESSION

BI814631

VERSION

BI814631.1 GI:15907251

KEYWORDS

EST.

SOURCE

malaria parasite P. falciparum.

ORGANISM

Plasmodium falciparum

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE

1 (bases 1 to 501)
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,
Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,
Bowers, Y., Gibbons, M., Rittner, E., Bennett, J., Jentes, E., Ronko, I.,
Tsagarashvili, R., Belaygorod, L., Franklin, C., Carr, L., Gow, A.,
Maguire, L., Richey, J., Madkins, J., Kennedy, S., Levinso, D.,
Waterston, R., Wilson, R., and Sibley, D.

TITLE

Unpublished (2001)

JOURNAL

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

COMMENT

Library was constructed by Deobam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 261.

FEATURES

source

1. 501
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site.1: EcoRI; Site.2:
XhoI; Library was constructed by Deobam Chakrabarti.
Total RNA samples were isolated from mixed stage
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyA+Tract
mRNA isolation system (Promega, WI) using streptavidin
magnisphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (3mg)
into EcoRI and XhoI sites of 1 Zapit vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the Exsist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

BASE COUNT 181 a 62 c 51 g 207 t

Query Match 3.3%; Score 161.2; DB 13; Length 501;
Best Local Similarity 72.7%; Pred. No. 6.5e-26;
Matches 208; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 4654 GACGGGAGAGAGTCAAGTCTCTGCACTTAAACAGAGAGATTAAGTGTGAG 4713
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 442
Db 501 GAAAGAGAGAGATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 442
QY 4714 AACCCAAACCTCTGCAATGAAAACATGGGGGTGTGACCCGATGCTAAATCACC 4773
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 382
Db 441 AATCCAAATCTCTGCAATGAAAACATGGGGGTGTGACCCGATGCTAAATC 382
QY 4774 GAGGAGACAGCGGCTCTTAACGAGAAAGAAATACATGCGAGTGTACTAAGCCGACTC 4833
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 322
Db 381 GAAAGAGATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTAATTA 322
QY 4834 TATCCACTCTGTGAGGAGATTTTCTGCTCCAGCTCAATTTCTGGGCACTTCCTCTCGT 4893
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 262
Db 321 TATCCACTCTGTGAGGAGATTTTCTGCTCCAGCTCAATTTCTGGGCACTTC 262
QY 4894 CTGATCTCATGCTGATCTGTACAGCTTCATCATTAATTAATTAATTAATTA 4939
11 TTTTAAAGACACATTAAGATCAAAATTTGATTCACCTTAATTAATTAATTA 216
Db 261 TTAATCTCATGCTGATCTGTACAGCTTCATCATTAATTAATTAATTAATTA 216

OY	4660	GAAGGTCGAATGTCTCCTCGAATACCAACAAGAAGGATTAAGTGGTGGAACAACCA	4719
Db	1	GAGAAATGTAATGTTTATTATAATTACAAACAAGAAGGTGATTAATGTGTTGAAAAATCCA	60
OY	4720	AACCCTACTGCATGAATAAACCAATGGCGGTGTGACGCCGATGCTTAATGACCCAGNNA	4779
Db	61	AATCCTACTGTTGAACGAAAATAATAGTGTGGATGTGATGACGATGCCACATGTACCGAAGA	120
OY	4780	GACAGCGGCTCTAACGGAAAGAAAATCACATGCGAGTGTACTTAAGCCGACTCTTATCCA	4839
Db	121	GATTGAGGTAGACGACGAAGAAGAAAATCATGATGTACTTAACACTGATTTCTTATCCA	180
OY	4840	CCTCTGACGGGATTTTTTGGCTCCAGGCTTAATTTCCGCGGATGTCCTCCTCGTGATG	4899
Db	181	CTTTTCGATGGATTTTTCGCGAGTTCTCTTCACTTCTTAGGAATATTCATTTCTTAATA	240
OY	4900	CTCATCTGATCCTCTACAGCTTCATCTTAATGATGATG	4939
Db	241	CTCATGTTAATTAATTATACAGTTTCATTTAAAAAATGTAG	280
RESULT 10			
LOCUS	BQ451036	455 bp	mRNA linear EST 29-MAY-2002
DEFINITION	PfESTabob0a10.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9DAI8 Q9DAI8 SURFACE PROTEIN-1 ;		
ACCESSION	BQ451036		
VERSION	BQ451036.1	GI:21254148	
KEYWORDS	EST.		
SOURCE	malaria parasite P. falciparum.		
ORGANISM	Plasmodium falciparum		
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	Tang,R., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Matra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagarelis,N.I., R., Belagorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Ritchey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D.		
TITLE	Mashu Plasmodium EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: L. David Sibley Mashu Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For Information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -40UP from Glibco High quality sequence stop: 426. Location/Qualifiers 1..455 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (Genehong, Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage asaponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the poly(A)-tract mRNA isolation system (Promega, WI) using streptavidin Magesphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb."		
FEATURES	Source		

BASE COUNT	179 a	46 c	56 g	174 t
ORIGIN				
Query Match	3.2%	Score 159:	DB 14:	Length 455:
Best Local Similarity	73.1%	Pred. No. 2e-25:		
Matches 204: Conservative	0:	Mismatches 75:	Indels 0:	Gaps 0:
QY 4661	AAGAGTGCAGTGTCTCTCTGACTACAAACAAGAGATAGTGCCTGGGAACCCAA	4720		
Db 1	AAGATGTAAATGTTTATTAAATTACAAACAAGAGGTGATATGTTGAAAAATCCAA	60		
QY 4721	ACCTACCTGCAGTAAACCAATGGCGGTGTGACGCCGCTGAATATGACCGAGAG	4780		
Db 61	ATTCCTACTTTGAACCAAAATATGGTGGATGTGATGCAGATGCCACATGTACCGAAGA	120		
QY 4781	ACACGGCTCTTACGGAAGAAATACATCGCATGTACTTACGACCCGACCTCTATCC	4840		
Db 121	ATTGACGTAGCAGCAGAAAGAAATACATGTGATGTACTTAAACCTGATCTTATCC	180		
QY 4841	TCTTCGACGGGATTTTTCCTCCAGCTCTAATTTCCCTGGGCAATCCTTCCTGGCATCC	4900		
Db 181	TTTTCGATGAGTAATTTCTGCACTTCCTTAACCTTTTGGAAATATCATCTTATTATATC	240		
QY 4901	TCATGCTGATCCCTGACAGCTTCAATCTAATGATGCATG	4939		
Db 241	TCATGTTAATATTATACAGTTTCATTTTAAAAATGTAG	279		
RESULT 11				
AU087573	500 bp	mrna	linear	EST 27-JAN-2001
LOCUS				
DEFINITION	AU087573 Sugano Malaria cDNA library Plasmidium falciparum 3D7 cDNA			
VERSION	clone XPrn5232 similar to P.falciparum gp190 (MSA1, MSP1, PMMSA)			
KEYWORDS	for precursor of major merozoite surface antigens, mRNA sequence.			
ACCESSION	AU087573			
VERSION	AU087573.1			
KEYWORDS	GI:12389714			
SOURCE	EST.			
ORGANISM	Plasmidium falciparum 3D7.			
REFERENCE	Plasmidium falciparum 3D7.			
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
TITLE	1 (bases 1 to 500)			
JOURNAL	Watanabe,J., Sasaki,M., Suzuki,Y., and Sugano,S.			
MEDLINE	PUBL-malaria: a database for a full-length enriched cDNA library			
COMMENT	from human malaria parasite, Plasmidium falciparum			
	Nucleic Acids Res. 29 (1), 70-71 (2001)			
	20574754			
	Contact: Junichi Watanabe			
	Institute of Medical Science			
	The University of Tokyo, Department of Parasitology			
	4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan			
	Tel: 81-3-5449-5378			
	Fax: 81-3-5449-5410			
	Email: jwatanbe@nagane.ims.u-tokyo.ac.jp			
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A., and Sugano			
	S., Construction and characterization of a full length-enriched			
	a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).			
FEATURES	Location/Qualifiers			
source	1..500			
	/organism="Plasmidium falciparum 3D7"			
	/strain="3D7"			
	/db_xref="taxon:36329"			
	/clone="XPrn5232"			
	/clone_id="Sugano Malaria cDNA library"			
	/dev_stage="erythrocytic stage"			
BASE COUNT	237 a	52 c	58 g	153 t
ORIGIN				
Query Match	3.2%	Score 158:	DB 9:	Length 500:
Best Local Similarity	59.3%	Pred. No. 3.5e-25:		

Matches : 288; Conservative : 0; Mismatches : 195; Indels : 3; Gaps : 1;

```
OY 1533 CTTGAGAGAAATGATGGAGAAACGGTACTCTTACAAAGTGGAGAACTGCACACCATAA 1592
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 CGTAGATAAATATTCATTCAGATGCAAGATATACATATATGTTGAAAACAAAGATATATTA 74
OY 1593 TACCTTTCATCCATGAGAAATTCATACATATCTTGAAGAGCTCACCAGAGCTTTAA 1652
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 TAAATTTTCATCTCTATTAATTCGTATATATGTCAAAAATTAATAAAGGCTCTTTC 134
OY 1653 GTATATGAGAGACTATCTTCGCGAGACATTTGTTGGAGAAAGCACTAAAGTATTACA 1712
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 ATATCTTAAAGATTTATCTTTTAAAGAAAGAAATTTCTGAAAAATTTTAAATCATATTA 194
OY 1713 GAATCTCAATAGTAAGATCGAAAAAGATGACAGCGCTGTTGGAACATTAAGAGA 1772
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 TACTTTGAAAACCTGGCCCTCGAAGCTGATATATAAAAAATTTACAGAGAAATTAAGATNG 254
OY 1773 TGAAGACAGTTGTTTGAAGAGATTTACAAAAGACAAAATAAACAGATGAGAGAT 1832
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 TGAAGACAAAATTCAGAAAAAAATTTTAAAGACTTAACACATTCAGCAAAATG--GTTG 311
OY 1893 CCTGAGAGTCTCCGATATTTGTTAAAGTCCAGAGTGCAGAGAGTGCCTCATGACAAAGT 1892
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 312 CTTAAGATATCTGATATTTGTTAAATTTACAGTACAAAAAGTTTATTAATTTAAAAAT 371
OY 1893 TGATGAACTCAAGAGACTCAACTCATCTCTGAAAGAGCTGAGTTAAACATATATACA 1952
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 372 AGAAGACTTAAAGAAAGATAGATTTTATTTTAAAAAAATGCACACACTAAAGATATGATCA 431
OY 1953 TGTGCGCAATAGTTATTAAGCAGAGAAATTAAGCAGAACCATATCTCATCTGATCA 2012
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 432 TGTACCAATATTTTATTAACCAAAATTAACAGAACCATATTTTAAATTTGATTA 491
OY 2013 GAAAGA 2018
      | | | | |
Db 492 AAAAGA 497
```

RESULT 12
AU088129 500 bp mRNA linear EST 27-JAN-2001
LOCUS AU088129 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
DEFINITION clone Xpfn6560 similar to P.falciparum (NF7) gene for mezozoite
surface antigen 1, mRNA sequence.
ACCESSION AU088129
VERSION AU088129.1 GI:12390270
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 500)
AUTHORS Matanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
JOURNAL MEDLINE
COMMENT Contact: Junichi Matanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jmatanabemange.jms.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
1..500
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/clone="Xpfn6560"

/clone_11b="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 197 a 65 c 58 g 144 t 36 others
ORIGIN

Query Match 3.2%; Score 156.4; DB 9; Length 500;
Best Local Similarity 57.6%; Pred. No. 8.1e-25;
Matches 250; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

```
OY 3161 ATCCGAACACCTACTGCAAGAACTTCTCAGTCTTCTTCAACAAGAAAGACCGAGA 3220
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 57 AACCCCATTAATGATTAACAAAACCTTCTGTTTCTTTAAACANTNAAAAAGAGTGAAA 116
OY 3221 TCGCCGAGACAGAGAAACACTGCGAAGACCAACAGATTCCTTCTCAACACTCAAGAGCC 3280
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 TAGCAGAAACATGANNANNACATTTAAAAAACCAAGANTATATTGAAACATTTAAAGGAC 176
OY 3281 TCGTCAAGTATTTAATATGCGAGTCTTCTCTGTAAGACTCTCCGAGAGAGCATCC 3340
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 TTGTAATAATATTTATATATGATGTAATCATCTCATTTAAAACTTAAGTAAATCAATTC 236
OY 3341 AGACCGAGATATACATACCGCACTCGAGAACTTCAAGTCTGTCTAAGCTCGAAGCA 3400
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 237 AAACGAAAGATATATATGCAATNTAGAAANNNTAGATATTAAGTNNAAATATGGA 296
OY 3401 AGCTGAAGGACAACTGGAACCTGAGAAAGAAAGACCTGACTCTTACGCGACTGC 3460
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 297 AACTCAATGATATTTATTTACATTTAGGAAGAAAAATTTATCTTCTTATCAAGTGCATTNC 356
OY 3461 ATCACCTGATCGCCGAGCTCAAGAGATGTCATTAAAGACAGAACTACACCGGCAATPAGCC 3520
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 357 ATCATTTAATTTACTGAATTAAGAGATTAATAAANTTAATTAATACAGGTNAATTCG 416
OY 3521 CAAGGAGAAATATATACAGACCGGATTAACGCCCTGGAATCTTACAGAAAGTCCGCGCTG 3580
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 417 CAAGTGAANATATTAAGAAAGTTTAACGAACTTTAAATCTTCAAAAAAATTTTTCNCA 476
OY 3581 AAGGACAGATGTC 3594
      | | | | |
Db 477 NAAGCAAAAGTTNC 490
```

RESULT 13
B1815239 303 bp mRNA linear EST 03-OCT-2001
LOCUS B1815239 PfEST0a16b11.y1 plasmodium falciparum 3D7 asexual cDNA Plasmodium
DEFINITION falciparum cDNA 5' similar to TR:Q90A18 Q90A18 SURFACE PROTEIN-1 ;
mRNA sequence.
ACCESSION B1815239
VERSION B1815239.1 GI:15908372
KEYWORDS EST.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 303)
AUTHORS Tang,R., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marrera,M., Hillier,L., Martin,J., Wylie,T., Dente,M., Theising,B.,
Bowers,Y., Gibbons,M., Riter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagarelisvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
TITLE MASHU Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University
putative full length read
vector to vector length is 433
Seq primer: -40up from Gibco
High quality sequence stop: 257.

FEATURES

source

Location/Qualifiers

1..303

/organism="Plasmodium falciparum"

/db.xref="taxon:5833"

/clone.lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab.host="DH10B (GeneHog, Invitrogen, Inc.)"

/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:

XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage

saponin(0.1%) lysed P. falciparum 3D7 infected

erythrocytes by the acidic guanidium-phenol chloroform

method. The poly A+ RNA was isolated by the polyA+ tract

mRNA isolation system (Promega, WI) using streptavidin

magnisphere particles. Directional cDNA libraries were

constructed by oligo d(T) priming of poly(A)+ RNA (5mg)

into EcoRI and XhoI sites of 1 ZapII vector using the Zap

cDNA synthesis kit (Stragene, CA). The average size of the

cDNA inserts in the library was between 1.0 and 1.5kb.

Clones were mass excised using the Exassist helper phage

(Stratagene), the phagemids were precipitated with PEG

8000 and extracted with phenol/chloroform. Phagemid DNA

was electroporated into DH10B cells."

BASE COUNT

106 a 41 c 41 g 115 t

ORIGIN

Query Match

Best Local Similarity 72.5%; Score 134.4; DB 13; Length 303;

Matches 174; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 4700 ATAGTCTGCGAGACCCAAACCTCTGCAATGATAAACAATGGCGGTGACGCCG 4759

DB 1 ATATAAGTGTGATAATCCAAATCCTACTGTGTACGAAATATATGTCGATGATGACG 60

QY 4760 ATGCTAATGACCGGAGAGACGCGCTTACGGAAGAAAGAAATACATGCGAGTGT 4819

DB 61 ATGCCCATGTACCGAAGAGATTCAGTAGCAGCAAGAAACATCATGTGATGT 120

QY 4820 CTAAAGCCGACCTCTGCACTCTGACGAGGATTTTCTGCTCCACTCAATTCCTG 4879

DB 121 CTAAACCTGATCTTATCCACTTTGATGATGTTTCTGCACTCTCACTTCTT 180

QY 4880 GCATCTCTCTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 4939

DB 181 GAATATCATCTTATTAATACATCATGTATATATATACAGTTTAAATAATGTACG 240

RESULT 14

BQ451367

LOCUS

DEFINITION BQ451367 419 bp mRNA linear EST 29-MAY-2002

Plasmodium falciparum 3D7 asexual cDNA Plasmodium

sequence. falciparum cDNA 5' similar to TR:Q9TYG8 Q9TYG8 MSAL PROTEIN ;, mRNA

ACCESSION

BQ451367

VERSION

BQ451367.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,

Marr, A., Hillier, L., Martin, J., Wylie, T., Dane, M., Theising, B.,

Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,

Tsagaris, V., Belaygorod, L., Franklin, C., Carr, L., Gow, A.,

Maguire, L., Richy, J., Wadkins, J., Kennedy, S., Levinso, D.,

Walter, R., Wilson, R., and Sibley, D.

TITLE

JOURNAL

COMMENT

Washu Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estelw@wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40up from Gibco.

FEATURES

source

Location/Qualifiers

1..419

/organism="Plasmodium falciparum"

/db.xref="taxon:5833"

/clone.lib="Plasmodium falciparum 3D7 asexual cDNA"

/lab.host="DH10B (GeneHog, Invitrogen, Inc.)"

/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:

XhoI; Library was constructed by Debopam Chakrabarti.

Total RNA samples were isolated from mixed stage

saponin(0.1%) lysed P. falciparum 3D7 infected

erythrocytes by the acidic guanidium-phenol chloroform

method. The poly A+ RNA was isolated by the polyA+ tract

mRNA isolation system (Promega, WI) using streptavidin

magnisphere particles. Directional cDNA libraries were

constructed by oligo d(T) priming of poly(A)+ RNA (5mg)

into EcoRI and XhoI sites of 1 ZapII vector using the Zap

cDNA synthesis kit (Stragene, CA). The average size of the

cDNA inserts in the library was between 1.0 and 1.5kb.

Clones were mass excised using the Exassist helper phage

(Stratagene), the phagemids were precipitated with PEG

8000 and extracted with phenol/chloroform. Phagemid DNA

was electroporated into DH10B cells."

BASE COUNT

189 a 49 c 59 g 122 t

ORIGIN

Query Match

Best Local Similarity 58.8%; Score 132.4; DB 14; Length 419;

Matches 248; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 4278 CCTGTACAAGACAGTACAGACAAATGCACTCTTCTGTAATTCACCTGGAGCCAAAGT 4337

DB 1 CTATACCAATTAAGTTAGTTAATAATGACGATTTACTTAATTAAGTAAAGCAAGAT 60

QY 4338 CCTCAACTATCTTACGAGACGCAATGTGAATTAATAAGAGCTGAATCACT 4397

DB 61 TAACGATTTGTAATGTTGAAGAAAGATGACGACATGTTAAATCACTTGTGATTT 120

QY 4398 CAAACAAATCCAGACAGCTGACATTTCAAGAAATTAACAATTTGCGGAATTCG 4457

DB 121 AAAGCAATTTGATGACAAATATAGATCTTTTAAAAACCTTACGACTTCGAACATTTAA 180

QY 4458 AGACCTGTCTACCGATTTATACCAACAATCTCTGACCAAGTTTCTGTCACCTGGCAT 4517

DB 181 AAATTTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 4518 GGTGTTGAAAACCTGCGCAAAACAGTGTGACATCTGCTGAGCGCACTGAGAGG 4577

DB 241 AG---TTCAAAATTTCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 297

QY 4578 CATGCTGAACATCTCCAGACACCAATGCGGAGAAAGATGCGGACATTAATGCGGCTG 4637

DB 298 TATGTTAAACATTTTCAACCAATGCGGATTAATTAATTAATTAATTAATTAATTAAT 357

QY 4638 TTTCAGGACATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4697

DB 358 TTTCAGGACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417

QY 4698 AG 4699

DB 418 TG 419

RESULT	15
LOCUS	B1814931
DEFINITION	B1814931 408 bp mRNA linear EST 03-OCT-2001 PFESTcaab0b05.y1 Plasmodium falciiparum 3D7 asexual cDNA plasmidium falciiparum cDNA 5' similar to TR:Q9UA18 Q9UA18 SURFACE PROTEIN-1 ; mRNA sequence.
ACCESSION	B1814931
VERSION	B1814931.1 GI:15907780
KEYWORDS	EST.
SOURCE	malaria parasite P. falciiparum.
ORGANISM	Plasmodium falciiparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 408)
AUTHORS	Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pepe,D., Marra,M., Hillier,L., Martin,J., Wyllie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,D., Jenkes,E., Ronko,I., Tsagarisilvill,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Macuire,L., Ritchey,J., Madkins,J., Kennedy,S., Levinso,D., Waterson,R., Wilson,R. and Sibley,D.
TITLE	WashU Plasmidium EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: L. David Sibley WashU Plasmidium EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University putative full length read vector to vector length is 409 Seq primer: -40UP from Gibco.
FEATURES	Location/Qualifiers
Source	1..408 /organism="Plasmodium falciiparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciiparum 3D7 asexual cDNA" /lab_host="DH10B (GeneHog, Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2 XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciiparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of l ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXassist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."
BASE COUNT	161 a 43 c 46 g 158 t
ORIGIN	
Query Match	2.5%; Score 124.4; DB 13; Length 408;
Best Local Similarity	72.5%; Pred. No.1.6e-17;
Matches 161; Conservative 0; Mismatches 61; Indels 0; Gaps 0;	
OY	4718 CAAACCCTACCGCATGGAACAATGGGGGTGGAGCCGATGCTAATGCACCGAGG 4777
DB	1 CAAATCCTACTTGTAAACCAATAATTAATGTTGGATGTGATGCACATGCCACATATTACCGAAG 60
OY	4778 AAGACGGGGCTCTAACGGAAAGAATACATGTGGAGTGCTAAGCCGACACTCTTAT 4837
DB	61 AAGATTACAGGTGACGACGAAGAAAATCACATGTGAATGTACTAAACCTGATTTCTATNC 120
OY	4838 CACTTCCTGACGGGGATTTTTTGTCTCAGCTTAATTTCTCTGGGCAATCTTCCTTCGTGA 4897

Db	121	CACCTTTGCATGCGTATTTCCTCAGTGTCTTAACCTCTTAGCAATATCATTTCTATTAA	180
OY	4898	TCCCTACTGCTGAATCCTGTACAGCTTCATCTAATAGATGATG	4939
Db	181	TACTCATGTTAATATTATACACTTTCATTAAAAATGAGG	222
RESULT	16		
LOCUS	BQ451572	418 bp	mRNA linear EST 29-MAY-2002
DEFINITION	PfSTcAbob11.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:O9UA18 O9UA18 SURFACE PROTEIN-1 ;		
VERSION	BQ451572		
KEYWORDS	mRNA sequence.		
SOURCE	BQ451572.1 GI:21254684		
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE	Plasmodium falciparum		
AUTHORS	1 (bases 1 to 418) Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Matta,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I., Tsagarisshvili,R., Belagorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Ritchey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D. Unpublished (2001)		
TITLE	Mashu Plasmodium EST Project		
JOURNAL	Contact: L. David Sibley Washu Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu		
COMMENT	Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -400P from Gibco. Location/Qualifiers 1..418 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (GeneHog, Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2 XhoI; library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage sporozoites (0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the poly(A)-tract mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of l ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the EXASist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."		
FEATURES			
SOURCE			
BASE COUNT	171 a 43 c 46 g 158 t		
ORIGIN			
Query Match	2.5%	Score 124.4;	DB 14; Length 418;
Best Local Similarity	72.5%;	Pred. No. 1.6e-17;	
Matches 161; Conservative	0; Mismatches 61;	Indels 0; Gaps 0;	
OY	4718	CAAACCACCTCGCAATAAAACATGGCGGTGTGACGCCGATGCTAATGACCGCAGG	4777
Db	1	CAAACTCCTACTTGTAAGCAAAAATAATGCTGGATGATGACATGATCCACATGTACCGAAG	60

QY	4778	AAGACAGAGGGCTCTAACGGAAAGAAATACATGTGGAGTCTTACGCCGCTCCCTATC	4837
Db	61	ANGATTTCAGGTGACGACGAGAAAGAAATACATGTGAATGTACTTAACCTGATTTCTTATC	120
QY	4838	CACCTTTCGACGGGATTTTTTTCCTCCAGCTCTAAATTTTCCTGGGCACTTCCTTCCTGCTGA	4897
Db	121	CACTTTTCGATGGATTTTTCCTCAGTTCTCTTAACCTTCTTAGAATATCATTTCTTATTA	180
QY	4898	TCCCTATGCTGATCTCTGTACAGCTTCATCTAATTAATGATG	4939
Db	181	TACTCATGTTAATATTAATACAGTTTCATTTAAAAAATGTAGG	222
RESULT 17			
LOCUS	B0451661	422 bp	mRNA linear EST 29-MAY-2002
DEFINITION	PfESTab07c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9VA18 Q9VA18 SURFACE PROTEIN-1 ;		
ACCESSION	B0451661		
VERSION	B0451661.1	GI:21254773	
KEYWORDS	EST.		
SOURCE	Plasmodium falciparum		
ORGANISM	malaria parasite P. falciparum.		
REFERENCE	Plasmodium falciparum		
AUTHORS	1 (bases 1 to 422) Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritzer, E., Bennett, D., Jentes, E., Ronko, I., Tsagarelis, V., R., Belagorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.		
TITLE	Washu Plasmodium EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: L. David Sibley Washu Plasmodium EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@washu.wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center for Information on obtaining a clone please contact: L. David Sibley (sibley@borcim.wustl.edu), Washington University Seq primer: -400P from Gibco.		
FEATURES			
Source	Location/Qualifiers 1..422 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /clone_lib="Plasmodium falciparum 3D7 asexual cDNA" /lab_host="DH10B (Genesig, Invitrogen, Inc.)" /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidium-phenol chloroform method. The poly A+ RNA was isolated by the polyA+-tract mRNA isolation system (Promega, WI) using streptavidin MagneSphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the Exsist1 helper phage (Stratagene), the phagemids were preclated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."		
BASE COUNT	175 a	43 c	46 g 158 t
Query Match	2.5%	Score:124.4	DB 14; Length 422;

[illegible]

Db 43 ATGATGAATATCCCTCTCAGAGATTTGAAATGATATGATATATTTAAACCTTAA 102
Oy 3877 GCAGCGCTATAGCTCTCTCAGAAACGCTGGAATAGTATGCTTCAATGTC 3936
Db 103 GCTGAGATATATGAACTTTAAAAACAAATTTGAAAAACATTTTCACTTAAATTTA 162
Oy 3937 AACGTGAAGACATTTCTGNAACGCCCTTTAATAGAGAGAAAATTTCAAGACCTTG 3996
Db 163 AATTGGAAGATATCTTAATTCAGCTCTTAAGAAAACGAAAATTTTCTTAAGATGATTA 222
Oy 3997 GAGACGACCTTGATTCCTATTAAGACCTGACCTCTCTTAAGTACGTTGTCAGACCCA 4056
Db 223 GAATCTGATTTAATGCAATTTAACAATATATCTCAAAATGATACATTTATGAAATTTCA 282
Oy 4057 TACAGTTCCTCAATGAAGAGAGAGGATTAATTTCTCTAGTATACATATATCAAG 4116
Db 283 TTTAAATATATGAAATTCGAAACAAAACACACTTTTAAAAAGTTACAAATATATATAA 342
Oy 4117 GACTCCATCGACACCGCATATCA 4139
Db 343 GAATCAGTAGAATAATGATATTTAA 365

RESULT 19
BM159597 597 bp mRNA linear EST 04-DEC-2001
LOCUS ESM562120 PyBS Plasmodium yoelii yoelii cDNA clone pICUG14 5' end,
DEFINITION mRNA sequence.
ACCESSION BM159597
VERSION BM159597.1 GI:17305278
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
REFERENCE 1 (bases 1 to 597)
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdivia,A.B.,
Fraser,C.M. and Carucci,D.J.
TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlone@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES
Location/Qualifiers
1..597
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="pICUG14"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByd mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dt)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt end. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

Query Match 2.3%; Score 112.2; DB 13; Length 597;
Best Local Similarity 52.4%; Pred. No. 1e-14;
Matches 292; Conservative 0; Mismatches 238; Indels 27; Gaps 1;

BASE COUNT 268 a 76 c 92 g 161 t
ORIGIN
Oy 487 AACATTCATGCGCTCAAAATATCTGATTGACGGTTACGAGAGATCAATGACCTGTAC 546
Db 36 AATGATATGAGTAAATATTTAATTTAAGAAAGAAAGAAATTTATGATTAATGATC 95
Oy 547 AAGTGAATTTTACTTGCACCTTGTGAAGGCCAAACTAAATGACCTTTGCCCAATGAC 606
Db 96 GCAATATAATTTTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 155
Oy 607 TATTCGCAATTCATTCATTTGAAATGACAGCCAAAGCTGACCTGATTTGAAGAAG 666
Db 156 TATTCGCAATTCATTCATTTGAAATGACAGCCAAAGCTGACCTGATTTGAAGAAG 215
Oy 667 TTGTCCTTCGATATCGAAGCCCTCTGACAAATCAATGACAGATGTGGAAAGATGAA 726
Db 216 GTAATTTTGTATGATTAAGAAACCAATAGAAATTTCAAGAGATTTGAAAGATGAA 275
Oy 727 GATTATATTTAAAGATATAGAACCATGCGAAGCAATTTAAGAGCTGTGCAAGATCC 786
Db 276 ATTTACATATGAAAGAAATTAAGAACTGTTCAGGCTTTAAAGCGCTTTTCTGAAAG 335
Oy 787 AAAAAGACCTA-----GACAAAATTAAGATGCAACC 819
Db 336 ACAAAAAAATATCAACCTTAAGCTAAGCAAGATTTGCAATGACGCTAGTTGTATGACCAT 395
Oy 820 AAGAGAGAGAAAGAGAAAGTTGTACAGGCCAGTACGACCTGTCCATATTAACAA 879
Db 396 AATATATATTAATAAAAAAACCAATATATACCAAGCTATGATTAATTTTAAACAAAA 455
Oy 880 CAGCTTGAAGAACCCATTAACCTCAACGCTGACGATGACGATGACGATGACGATGAC 939
Db 456 CAATTAGCTGAAATCAAAAGGTTGCGAAGCTTTGAAAGAAAGAGCTTTCTCATTTAAAG 515
Oy 940 AAGAATGAATAATATCAAAAGCTGCTGCAAGATTAATGAATTAAGATCTCCGCCA 999
Db 516 AAAAATGATGCATCAACACATTAATGCGAACAATTTGAAGTTCTCAATCTGCCCCGTC 575
Oy 1000 GCCAATCTGGGACAC 1016
Db 576 GTCACTGCCGGAACAC 592

RESULT 20
A0086246 500 bp mRNA linear EST 27-JAN-2001
LOCUS A0086246
DEFINITION A0086246 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
clone xPn2175 similar to P.falciparum gp190 (MSA1, PMSA1)
for precursor of major merozoite surface antigens, mRNA sequence.
ACCESSION A0086246.1 GI:12388387
VERSION A0086246.1
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7.
REFERENCE 1 (bases 1 to 500)
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
JOURNAL 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@manage.ims.u-tokyo.ac.jp
Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano
S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES

source

Location/Qualifiers
1..500
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/clone="ypn2175"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 244 a 69 c 72 g 115 t
ORIGIN

Query Match 2.2%; Score 108.8; DB 9; Length 500;
Best Local Similarity 58.2%; Pred. No. 6e-14; Mismatches 137; Indels 0; Gaps 0;
Matches 191; Conservative 0;

1745 AGACGCTGTGGACATTAGAGAGATGAAGACAGTTGTTGAGAAAGATTACAA 1804
19 ATTCATATATTACAGAGAAATTAAGAGAGTGAACAAATTTCTAGAAAAATTTA 78
1805 AAGACGAAATTAACAGATGAGAGATCCCTGGAGTCTCCGATTTGTTAAAGTCCAG 1864
79 AAGGACTAACACATGACGAAATGTTCCCTTAGAGATCTGATTTGTTAAATTCAG 138
1865 TGGAGAGGTGCTCTCATTGAGACAGATGATGACTCAAGACTCACTCATCTGA 1924
139 TACAAAAGCTTTTATTTATTTAAATAAGAACTTAGAAAGATGAAATTTTAA 198
1925 AGAACCTGGAGTTAAACATATATATGATGCGGATATGTTTAAGCAGAGATTAAGC 1984
199 AAATGACACACATAAAGATAGTATCTACCAATATTTAAACCCAAATAAAC 258
1985 AGGACCATACTACTCTGATCTGATCAAGAGATAGCAAAAGTTCATGC 2044
259 CAGAACCATATTTATTTATTTATTTAAAGAAAGATGATTAATTAAGATTTATAC 318
2045 CCAAGTCGAGAGCTGATCAAGAGA 2072
319 CAAAAGTAAAGACATGTTAAAGAGAGA 346

RESULT 21
BM159731 755 bp mRNA linear EST 04-DEC-2001
LOCUS EST1622254 PyBS Plasmodium yoelii yoelii cDNA clone pYCUH95 5' end,
DEFINITION mRNA sequence.

ACCESSION BM159731
VERSION BM159731
KEYWORDS
SOURCE EST
ORGANISM Plasmodium yoelii yoelii.
Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 755)
AUTHORS Carlson, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
Fraser, C.M. and Carucci, D.J.
Plasmodium yoelii EST project at TIGR

TITLE Unpublished (2001)
JOURNAL Contact: Jane Carlton
COMMENT Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

*Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.

FEATURES

source

Location/Qualifiers

1..755
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCJH95"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with PY17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT 323 a 103 c 136 g 193 t
ORIGIN

Query Match 2.2%; Score 108.8; DB 13; Length 755;
Best Local Similarity 52.4%; Pred. No. 6.6e-14;
Matches 285; Conservative 0; Mismatches 232; Indels 27; Gaps 1;

487 AACATTCATGCTTCAATATCTGATGAGCTTAGCAGATCAATGAATCCTGTAC 546
39 AATAGATATGAGTAAATTTTAAATTAAGAAAGTAAAGAAATTTATGATTAATGAC 98
547 AACTGAAATTTCTACTTCGACTTGCCTGATAGGCGCAAACTGATACGTTGGCCATGAC 606
99 GCAATTAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 158
607 TATTGCAATTTTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
159 TATTGCAATTTTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 218
667 TTGCTTTCGATATGCAAGCCCTCTGACACATCAAGACATGTTGGAAAGATGAA 726
219 GTATTTTTCGATATGCAAGCCCTCTGACACATCAAGACATGTTGGAAAGATGAA 278
727 GATTATATTTAAAGATTAAGAGACCATGGAACATTTAGCAGCTGATCGAAGATCC 786
279 ATTACATAGCAAAATTAAGAAAGCTGTGACGTTTAAACCTCTTATTTGCTGAAGAA 338
787 AAAAGACCATTA-----GACAAAATTAAGATGCAAC 819
339 ACAAAATAATTAACACCTGAGGTAAAGAGATGCAATGATGATGATGATGATGAT 398
820 AAGGAG 879
399 AATATATATTAAG 458
880 CAGCTTGAAG 939
459 CAATTAGCTGAATTAAG 518
940 AAGATGAATATTCAG 999
519 AAAAATGATGCCATCAACCATTTATGCGACACAAATTAAGATGCTCAATGCTGCCCGCTC 578
1000 GCCA 1003
579 GTCA 582

RESULT 22
 LOCUS BM160423 753 bp mRNA linear EST 04-DEC-2001
 DEFINITION EST562946 PyBS Plasmodium yoelii yoelii cDNA clone pYCJR11 5' end,
 mRNA sequence.
 ACCESSION BM160423
 VERSION BM160423.1 GI:17306104
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 753)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdivya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlone@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES
 source
 1..753
 Location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJR11"
 /clone_lib="PyBS"
 /dev_stage="asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/CByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dt)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (pad-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT 335 a 101 c 113 g 204 t
 ORIGIN

Query Match 2.2%; Score 107.6; DB 13; Length 753;
 Best Local Similarity 58.9%; Pred. No. 1.2e-13;
 Matches 185; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 487 AACATTCAGTGGCTCAATATCTGATGACGGTTAGCAAGATCAATGACTCCTGAC 546
 ||| |||| ||||| ||||| || ||||| || ||||| || ||||| || |||||
 DB 172 AATAGATATGAGTAATAATATTATGAAAGTAAGAAATTAATGATTAATGCAC 231
 QY 547 AAGTTGATTTCTACTTGACTGCTAAGGGCCAAAGCAATGACGTTGCCCAATGAC 606
 ||||| ||| || || || || || || || || || || || || || || || || ||
 DB 232 GCAATAAATTTTATATGATGATCTAGAGTAATAATTAATGATATGTGCAAAATAAT 291
 QY 607 TATGTCAATTCATTCATTCATTTGAAGATCAGCAACGAGTTGAGCTATGGAAG 666
 ||||| ||||| || || || || || || || || || || || || || || || || ||
 DB 292 TATTTGAAATTCCTGAAACATCTTAATAATTAGTGAAAGAAAGCAATGCTTAAAAA 351

QY 667 TTGCTTCGATATGCGAAGCGCTCTGACATCATCAAGACAAATGCGAAGATGAA 726
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 DB 352 GTAATTTTGGCTTATGAAACCAATAGAAATATTTCAGACGATTTAAAAAGTTAGAA 411
 QY 727 GATTATTTTAAAGATTAAGAACCATCGAATCATTTACGAGCTGATCGAAGATCC 786
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 DB 412 ATTTCATGAAAGAAATTAAGAACTGTTGAGCTTTAAACGCTCTTATTCGTGAAGAA 471
 QY 787 AAAAGACCATAGA 800
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 DB 472 ACAAATAAATAACA 485

RESULT 23
 LOCUS BM165725 678 bp mRNA linear EST 04-DEC-2001
 DEFINITION EST568248 PyBS Plasmodium yoelii yoelii cDNA clone pYCMR8 5' end,
 mRNA sequence.
 ACCESSION BM165725
 VERSION BM165725.1 GI:17298957
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Plasmodium yoelii yoelii.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 678)
 AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdivya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished (2001)
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlone@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES
 source
 1..678
 Location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCMR78"
 /clone_lib="PyBS"
 /dev_stage="asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/CByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dt)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (pad-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT 291 a 93 c 120 g 174 t
 ORIGIN

Query Match 2.1%; Score 105; DB 13; Length 678;
 Best Local Similarity 52.4%; Pred. No. 4.7e-13;
 Matches 277; Conservative 0; Mismatches 225; Indels 27; Gaps 1;

FEATURES

Location/Qualifiers
1. 647

/organism="Plasmodium berghei"

/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"

/db_xref="taxon:3821"

/clone_lib="Pb MBN #21"

/dev_stage="asexual blood forms"

/lab_host="Mus musculus"

/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda Zap, Site_1: EcoRV, Site_2: EcoRV;

Genomic DNA was prepared from asynchronous blood stage

forms of the cloned ANKA isolate of P. berghei grown in

laboratory Swiss white mice. The DNA was purified from

contaminating host DNA by Hoechst Dye 33258-CsCl

ultracentrifugation and precipitated. Purified DNA was

digested with mung bean nuclease in the presence of 36-38%

formamide at 50 C, as described (Vernick, K.D., Imberski,

R.B., and McCutchan, T.F., 1988. Nucleic Acids Research

16:6883-6896). The ends of the digestion fragments were

polished using T4 DNA polymerase, and the fragments size

selected in the range 500-2000 bp. These were ligated into

the EcoRV-cleaved and dephosphorylated pBluescript SK(+)

vector. Recombinant plasmids were used to transform E.

coli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

FEATURES

source

Location/Qualifiers

1. 660

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone_lib="PyB5"

/dev_stage="asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was

collected from BALB/cByJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

terminals were treated with Pfu DNA polymerase and EcoRI

adaptors. Ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HybridZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HybridZAP vector and plasmid DNA

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

isolated."

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

Seq primer: ADP.

FEATURES

Location/Qualifiers

1. 647

/organism="ANKA clone 15cyl (clone of the ANKA 8417 clone)"

/db_xref="taxon:3821"

/clone_lib="Pb MBN #21"

/dev_stage="asexual blood forms"

/lab_host="Mus musculus"

/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda Zap, Site_1: EcoRV, Site_2: EcoRV;

Genomic DNA was prepared from asynchronous blood stage

forms of the cloned ANKA isolate of P. berghei grown in

laboratory Swiss white mice. The DNA was purified from

contaminating host DNA by Hoechst Dye 33258-CsCl

ultracentrifugation and precipitated. Purified DNA was

digested with mung bean nuclease in the presence of 36-38%

formamide at 50 C, as described (Vernick, K.D., Imberski,

R.B., and McCutchan, T.F., 1988. Nucleic Acids Research

16:6883-6896). The ends of the digestion fragments were

polished using T4 DNA polymerase, and the fragments size

selected in the range 500-2000 bp. These were ligated into

the EcoRV-cleaved and dephosphorylated pBluescript SK(+)

vector. Recombinant plasmids were used to transform E.

coli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

FEATURES

Location/Qualifiers

1. 660

/organism="ANKA clone 15cyl (clone of the ANKA 8417 clone)"

/db_xref="taxon:3821"

/clone_lib="Pb MBN #21"

/dev_stage="asexual blood forms"

/lab_host="Mus musculus"

/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda Zap, Site_1: EcoRV, Site_2: EcoRV;

Genomic DNA was prepared from asynchronous blood stage

forms of the cloned ANKA isolate of P. berghei grown in

laboratory Swiss white mice. The DNA was purified from

contaminating host DNA by Hoechst Dye 33258-CsCl

ultracentrifugation and precipitated. Purified DNA was

digested with mung bean nuclease in the presence of 36-38%

formamide at 50 C, as described (Vernick, K.D., Imberski,

R.B., and McCutchan, T.F., 1988. Nucleic Acids Research

16:6883-6896). The ends of the digestion fragments were

polished using T4 DNA polymerase, and the fragments size

selected in the range 500-2000 bp. These were ligated into

the EcoRV-cleaved and dephosphorylated pBluescript SK(+)

vector. Recombinant plasmids were used to transform E.

coli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

colli XL10-Gold host cells."

FEATURES

Location/Qualifiers

1. 660

/organism="ANKA clone 15cyl (clone of the ANKA 8417 clone)"

/db_xref="taxon:3821"

/clone_lib="Pb MBN #21"

/dev_stage="asexual blood forms"

/lab_host="Mus musculus"

/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda Zap, Site_1: EcoRV, Site_2: EcoRV;

Genomic DNA was prepared from asynchronous blood stage

forms of the cloned ANKA isolate of P. berghei grown in

laboratory Swiss white mice. The DNA was purified from

contaminating host DNA by Hoechst Dye 33258-CsCl

ultracentrifugation and precipitated. Purified DNA was

digested with mung bean nuclease in the presence of 36-38%

formamide at 50 C, as described (Vernick, K.D., Imberski,

R.B., and McCutchan, T.F., 1988. Nucleic Acids Research

16:6883-6896). The ends of the digestion fragments were

polished using T4 DNA polymerase,

ORGANISM Plasmodium yoelii yoelii

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 777)

AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,

Fraser, C.M. and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR

JOURNAL Unpublished (2001)

COMMENT Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference

Reagent Resource Center, ATCC

http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADP.

FEATURES

source

1..777

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCQNZ4"

/clone_id="PYBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4. At 20-25% parasitemia, blood was

collected from BALB/cBYJ mice infected with Py17XL

parasites, and leukocytes removed by passage over

microcrystalline cellulose columns. Total RNA was

isolated using the guanidinium isothiocyanate method, and

mRNA isolated using oligo(dT)-cellulose chromatography.

First strand cDNA synthesis was completed using a 50-base

primer and reverse transcriptase in the presence of

5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI

adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column.

Size-fractionated cDNA was precipitated and ligated to

HydriZAP arms directionally using EcoRI-XhoI cleaved arms.

After packaging, the phagemid vector (PAD-GAL4) was

excised from the HydriZAP vector and plasmid DNA

isolated."

BASE COUNT 328 a 124 c 124 g 201 t

ORIGIN

Query Match 1.8%; Score 91.2; DB 13; Length 777;

Best Local Similarity 47.5%; Pred. No. 6.9e-10;

Matches 302; Conservative 0; Mismatches 333; Indels 1; Gaps 1;

OY 2965 AAACGTAAAGAGCTGGAGAGGACATCAATTAAGCTGAAGAAGACATGCAAGCTTC 3024

DB 3 AAAGTTAT 62

OY 3025 GACCTGTACAAACAGTCAAACTGAGAGACTCTTCGACAAAGAGACAGACATC 3084

DB 63 GATCGTATATTTACATATTAATTAATTAAGAAAGATTATTAACAACATGACAAATTT 122

OY 3085 GCGAAGTATAGATGAGATGAGAGTGAAGTCTGCTCAAGAGACAGCTTGAAGCAAA 3144

DB 123 CAACATACCAATGACAAATTAAGAGATCTTGTATATTGAAGACGATTTTAAAGA 182

OY 3145 CTCAACTCAGCAGCAATCCGAAACAGCTGACAGAACTTCTCAAGTCTTCAACAAG 3204

DB 183 AAACAACCTTTAATGCGATATTTATATATTAATGTTATGTAATTTCTTTAACAAG 242

OY 3205 AAGAAGAGCGGATGCGCGAGAGAGAACTGCGAGACACCAAGATCTCTC 3264

DB 243 AGAAGAGAGAGCTGAAAAACATATGTAGATTAATGCAATTAATAAAATACGATATGTTATTA 302

OY 3265 AAACATGACAAAGCGCTGCTCAAGTATTAATGCGAGATCTCTCTGAGAGACTCTC 3324

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

DB 303 AAATACTACAAAGCTGCTACTTAATATTTTACTTCTGAAGCTGTTCTTTAAAAATTA 362

OY 3325 TCCGAGAGAGCATCCAGACCGAGATATACAGCCACCTCGAAGCTTCAAGTCTCG 3384

DB 363 TCTTAAGATCATCTTGACAGAGAAATTCATTTTGGAAAAATCGAAAAATTCGACATAC 422

OY 3385 TCTAAGCTCGAAGGCAAGCTGAGACCAAGCTGAGAGAGAGAGAGAGCTGAGTAC 3444

DB 423 AGTCATTAAGATTAAGATTAATAAAAAATTTATTTAGGAAGAGAAAGAAATTTTCATAT 482

OY 3445 CTCTCTAGCGGACATCATCCTGATCGCCGAGCTCAAGAGAGTATTAAGACAGAAC 3504

DB 483 GTATCAGAGGTTTACACACGATTTTGAAGATTTAAAGAACTTATTAAGATTAAGAC 542

OY 3505 TACACCGGCAATAGCCCAAGCAGAAATTAATACAGACGTAATACGACTGGAATCTTAC 3564

DB 543 TATACCGGAAAAAAAACCTGATATATGTCCTCGTAAGT-ACCAATGATTCGAAACAAATAT 601

OY 3565 AAGAAGTCTGCTGCGAAGAAACAGATGTCGCCACT 3600

DB 602 AAAGATTTGCTTCCAAAGGAGTACAGATTTCACCT 637

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

III IIIIIIIII III IIIII III III IIIII III III

	source	1.	362	/organism="Plasmodium falciparum"
		/db_xref="taxon:5833"		
		/clone_id="Plasmodium falciparum 3D7 asexual cDNA"		
		/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"		
		/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyA-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of l ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were preclipped with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."		
	BASE COUNT	153 a	32 c	33 g 144 t
	ORIGIN			
	Query Match	1.7%	Score 86.2;	DB 14; Length 362;
	Best Local Similarity	70.6%;	Pred. No. 8e-09;	
	Matches 115; Conservative	0;	Mismatches 48;	Indels 0; Gaps 0;
Oy	4777 GAAGCAGCGGGCTTAACGGAAGAANAATTCACATCGAGTGTACTAACCGGACATCAT	4836		
Dd	2 GAAATTACAGTGACGACGAGAAAGAAATCAATGTGAATGACTAAACCTGATTTCTAT	61		
Oy	4837 CCACCTCTTCGAGCGGGATTTTTTGCCTCCAGCTCTAATTTCTCGGCGATCCTCTCGCTG	4896		
Dd	62 CCACCTTTTCGATGGATTTTCTGCGAGTTCCTCTAAGGAAATTCATTTCTATTATA	121		
Oy	4897 ATCCATCATGCTGATTCCTGTACAGCTTCATCTTAATGATCGATG	4939		
Dd	122 ATACTCATGTTAATTAATTATACAGTTTCATTTAAAAAATGTAGG	164		
RESULT 33				
LOCUS	B0577302	352 bp	mRNA	linear EST 19-JUN-2002
DEFINITION	PfPSToaab17c05.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum cDNA 5' similar to TR:Q9VA18 Q9VA18 SURFACE PROTEIN-1 ; , mRNA sequence.			
ACCESSION	B0577302			
VERSION	B0577302.1	GI:21480619		
KEYWORDS	EST.			
SOURCE	Plasmodium falciparum.			
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
REFERENCE	1 (bases 1 to 352)			
AUTHORS	Tang,R., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennettt,C., Jentes,E., Ronko,I., Tesgareishvili,L.R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Meuterie,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D.			
TITLE	WashU Plasmodium EST Project			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: L. David Sibley WashU Plasmodium EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: L. David Sibley (lsibley@borcim.wustl.edu), Washington University			

FEATURES	Seq primer: -400p from GlbCO.
Source	Location/Qualifiers
	1..352
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone_lib="plasmodium falciparum 3D7 asexual cDNA"
	/lab_host="DH10B (Genesig, Invitrogen, Inc.)"
	/note="Vector: pBluescript SK plus; Site_1: EcoRI. Site_2: XhoI. Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-trac mRNA isolation system (Promega, WI) using streptavidin Magnetic particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 zapII vector using the zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were preclipped with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."
BASE COUNT	143 a 33 c 32 g 143 t 1 others
ORIGIN	
Query Match	1.7%; Score 85.2; DB 14; Length 352;
Best Local Similarity	70.4%; Pred. No. 1.4e-08;
Matches 114; Conservative	0; Mismatches 48; Indels 0; Gaps 0;
OY 4778	AAGACAGCGGCTCTTAACGGAAGAAATACATGCGAGTCTAGACCCGACTCTATGC 4837
Db 1	AAGATTGAGTACGACGACGAGAAAGAAATACATGATGATGATCAAACTATTCTTATC 60
OY 4838	CACCTCTGAGGGGATTTTTTGGCTCCACGCTCAATTCCTGGGACATCTCTCCGCGGA 4897
Db 61	CACCTTTCGATGATGATTTTTCGACAGTCTCTCACTTTTGAAGATATCATTCATTATTA 120
OY 4898	TCCCTCATGCTGATCTCTGACAGCTTCATCTAATAGATCGATG 4939
Db 121	TACCTATGTTAATATTATACAGTTTCATTATTAATAAATGTAG 162
RESULT 34	
LOCUS	T18003 282 bp mRNA linear EST 30-AUG-1994
DEFINITION	0325c3 cbSPFHB3.1, Debopam Chakrabarti Plasmodium falciparum cDNA
ACCESSION	T18003
VERSION	T18003.1
KEYWORDS	EST.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 282)
TITLE	Chakrabarti, D., Reddy, G.R., Dame, J.B., Almtira, E.C., Lalpiz, P.J.,
JOURNAL	Fel, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M.
MEDLINE	Analysis of Expressed Sequence Tags from Plasmodium falciparum
COMMENT	Mol. Biochem. Parasitol. 66, 97-104 (1994)
	95075403
	Contact: Debopam Chakrabarti
	Department of Molecular Biology and Microbiology
	University of Central Florida
	Orlando, FL 32816-2360
	Tel: 407 384 2061
	Fax: 407 384 3095
	Email: dchakrepegasus.cc.ucf.edu
	Seq primer: T3.
FEATURES	
source	Location/Qualifiers
	1..282
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone="0325c"

FEATURES	Seq primer: -400p from GlbCO.
Source	Location/Qualifiers
	1..352
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone_lib="plasmodium falciparum 3D7 asexual cDNA"
	/lab_host="DH10B (Genesig, Invitrogen, Inc.)"
	/note="Vector: pBluescript SK plus; Site_1: EcoRI. Site_2: XhoI. Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-trac mRNA isolation system (Promega, WI) using streptavidin Magnetic particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 zapII vector using the zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were preclipped with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."
BASE COUNT	143 a 33 c 32 g 143 t 1 others
ORIGIN	
Query Match	1.7%; Score 85.2; DB 14; Length 352;
Best Local Similarity	70.4%; Pred. No. 1.4e-08;
Matches 114; Conservative	0; Mismatches 48; Indels 0; Gaps 0;
OY 4778	AAGACAGCGGCTCTTAACGGAAGAAATACATGCGAGTCTAGACCCGACTCTATGC 4837
Db 1	AAGATTGAGTACGACGACGAGAAAGAAATACATGATGATGATCAAACTATTCTTATC 60
OY 4838	CACCTCTGAGGGGATTTTTTGGCTCCACGCTCAATTCCTGGGACATCTCTCCGCGGA 4897
Db 61	CACCTTTCGATGATGATTTTTCGACAGTCTCTCACTTTTGAAGATATCATTCATTATTA 120
OY 4898	TCCCTCATGCTGATCTCTGACAGCTTCATCTAATAGATCGATG 4939
Db 121	TACCTATGTTAATATTATACAGTTTCATTATTAATAAATGTAG 162
RESULT 34	
LOCUS	T18003 282 bp mRNA linear EST 30-AUG-1994
DEFINITION	0325c3 cbSPFHB3.1, Debopam Chakrabarti Plasmodium falciparum cDNA
ACCESSION	T18003
VERSION	T18003.1
KEYWORDS	EST.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota, Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 282)
TITLE	Chakrabarti, D., Reddy, G.R., Dame, J.B., Almtira, E.C., Lalpiz, P.J.,
JOURNAL	Fel, R.J., Yang, T.P., Rowe, T.C. and Schuster, S.M.
MEDLINE	Analysis of Expressed Sequence Tags from Plasmodium falciparum
COMMENT	Mol. Biochem. Parasitol. 66, 97-104 (1994)
	95075403
	Contact: Debopam Chakrabarti
	Department of Molecular Biology and Microbiology
	University of Central Florida
	Orlando, FL 32816-2360
	Tel: 407 384 2061
	Fax: 407 384 3095
	Email: dchakrepegasus.cc.ucf.edu
	Seq primer: T3.
FEATURES	
source	Location/Qualifiers
	1..282
	/organism="Plasmodium falciparum"
	/db_xref="taxon:5833"
	/clone="0325c"

OY	953	TCAAGACATGCTGCACAAGATTATTAATGAATTAAGCATTCGCCGACGCA	1003
Dd	421	TCAATCATTATGGCACAAATTGATTTCTCATATGCTCCCGCTGTCA	471
RESULT	38		
LOCUS	BM166688		
DEFINITION	BM166688	769 bp mRNA linear EST 04-DEC-2001	
ACCESSION	EST5692211	PYBS Plasmodium yoelii yoelii cDNA clone pycnv80 5' end,	
VERSION	BM166688	mRNA sequence.	
KEYWORDS	BM166688.1	GI:17299920	
SOURCE	EST.		
ORGANISM	Plasmodium yoelii yoelii.		
REFERENCE	Plasmodium yoelii yoelii		
AUTHORS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
	1 (bases 1 to 769)		
	Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdya,A.B.,		
	Frasser,C.M. and Carucci,D.J.		
	Plasmodium yoelii EST project at TIGR		
	unpublished (2001)		
TITLE	Contact: Jane Carlton		
JOURNAL	Parasite Genomics Group		
COMMENT	The Institute for Genomic Research		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-530-9319		
	Fax: 301-838-0208		
	Email: carlton@tigr.org		
	For clone info, please contact the Malaria Research and Reference		
	Reagent Resource Center, ATCC		
	http://www.malaria.mr4.org/pages/index.html		
	Seq primer: ADF.		

FEATURES	Location/Qualifiers
source	1. .769

/organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCNB8"
 /clone_1lb="PYBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: pAD-GAL4: At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with Pyl17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dt)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to Hybridzap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the Hybridzap vector and plasmid DNA isolated."

BASE COUNT	340 a	101 c	104 g	224 t
ORIGIN				
Query Match	1.6%	Score 79;	DB 13;	Length 769;
Best Local Similarity	46.6%	Pred. No. 4.2e-07;		
Matches 285; Conservative	0;	Mismatches 325;	Indels 1;	Gaps 1;

QY	2908	TTTCACTTAACCTTCGTAAGATCTTAAGCCGATGATATCAACTCTCTTAAGCATGAATCTCTAAA	2967
Db	111	TTTCATTAATTTTATTAATAAATCTTAAAAAGAGATTAAATTAACAAGCATCTGACACCTGGAAAAA	170
QY	2968	CCTAAGAAGCTGGAAGAGACATCATATAGCTGAAGAAGACACTGCACCTGAGCTTCGCAC	3027
Db	171	GTTAATTCATTATTAATCTTGAAGATTCCTCACTTAATAAGGAATTATTCAGAACATTATATATGAT	230

OY	3028	CTGACCAACGAGTCAAACTGAACTGGAGAGACTCTTCGACAGAAAGAACAGTGC	3087
Db	231	CGTATTCTACATATTAATTAATAANTTAGAAAGATTATATTACCAACATGAACTTA	290
OY	3088	AAGTATTAAGATGACAGATCAGAGAGTTGACTCTGCTCAAGAGACGCTGAAAGCAAACTC	3147
Db	291	CTAACCATCCAGCAAAATTAGATCTTGTATATTGAAAGCACGATTTATTAAGAAAA	350
OY	3148	AACTCACAGCAATCCGAAACAGCTACTGGAGAACTTCAGTGTCTTCACAGAAAG	3207
Db	351	CAAACTCTTAATGGCGATTTTATTATATTAAATGTTATGTAAATTTCTTTAACAGAGA	410
OY	3208	AAGGAAGCCGAGATGCGCCGAGACAGAGAACACTCTGGAACACCAAGATTCCTCTCAA	3267
Db	411	AGAGAAAGCTGAAAAACAATATGTAGATATCATTTAAAAAATACGATATCTTTATAAAA	470
OY	3268	CACGTAAAGGCGTCGTCAAATTTATTAATGCGCGATCTTCCTGTCAACACTCTCC	3327
Db	471	TACTACAAAGCTCGTACTAAATATTTTACTTCTGAGCTGTTCCTTTAAAAACATTATCT	530
OY	3328	GAGGAGAGCATCCAGACCGAGGATTAACACGCCACGCCCTCGAAGACTTCAGAGTCTGCT	3387
Db	531	AAAGATCATCTTGACAGGAATCCAAATTAATTGAAAAATCGAAAAATTCAGAGCATACGT	590
OY	3388	AAGCTCGAAGCAAGCTGAAGACAACTGTAACTGGAGAAAGAAAGAACTCAGCTACC	3446
Db	591	CGATTGAGATTAAAGTTTAAAAAAAATTTATTAATTGAGAAAGAAAGAAATTTCAATGCT	650
OY	3447	CTTAGCGGACTGCATCCATCGATGCGCGAGACTCAAGGAAGTCAATTAAAGACAGAACTA	3506
Db	651	ATCAGGAGGTTTACACCCAGCTATTGGAAGATTTAAAGAACTTATTTAAAAAGATTAAGACTA	710
OY	3507	CACCGGCAATA	3517
Db	711	TACCGGAAAAA	721

	RESULT	39		
	B13042/c			
	LOCUS		1147 bp	DNA linear GSS 14-MAY-1997
	DEFINITION	B13042 T3OM24-Sp6.1 TAMU Arabidopsis thaliana genomic clone T3OM24,		DNA sequence.
	ACCESSION	B13042		
	VERSION	B13042.1 GI:2094174		GSS.
	KEYWORDS	tale cress.		
	SOURCE	Arabidopsis thaliana		
	ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytia; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 1147)		Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.
REFERENCE	AUTHORS			
TITLE	JOURNAL COMMENT			
	BAC end sequences at ATGC Unpublished (1997) Other_GSSs: T3OM24-Sp6, T3OM24-Sp6.2			

Contact: Ecker, J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu

```

FEATURES
  source
    Seq primer: 5pb
    Class: BAC ends
    High quality sequence start: 289
    High quality sequence stop: 292.
    Location/Qualifiers
      1..1147
        /organism="Arabidopsis thaliana"
        /strain="Columbia"
        /db_xref="taxon:3702"

```



```

      /clone="T30M24"
      /sex="hermaphrodite"
      /note="Vector: BelopacII; Site_1: HindIII; Site_2: HindIII
      ; Produced by Rod Wing"
BASE COUNT      3 a      4 c      46 g      553 t      541 others
ORIGIN

```

```

Query Match      1.6%: Score 78.8; DB 17; Length 1147;
Best Local Similarity 23.8%; Pred. No. 5.2e-07;
Matches 251; Conservative 0; Mismatches 805; Indels 0; Gaps 0;

Qy 695 ACAACATCAAGACATGTTGGAAAGATGAAATATTATTTAAAGATAAGACCA 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1145 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1086

Qy 755 TCGAGACATTACGAGCTGATCGAAGATCCAAAAGACCATAGACAAAATAAGATG 814
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1026

Qy 815 CAACCAAGAGGAGAAAGAAAGAGTGTACCGCCAGCAGACCTGTCCATCTATA 874
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 966

Qy 875 ACAACACCTTGAAGAGCCCATTAACCTCATCGCCTACTGGAAGCGCATAGACACC 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 965 NANANNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNNANNA 906

Qy 935 TCAGAGAGATGAAATATTCAGAGACCTGCTGCAGACATTTATGAAATTAAGATCCTC 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 905 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 846

Qy 995 CGCCGACACTCTGGAACACCCCTTAACAGCTCTGAGAGAGAAAGATAGAGG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 845 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 786

Qy 1055 AGCAGAGAAAGAGATCAAGAGATGCCAAACCATTAAGTCAACATAGATTCTCT 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 785 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 726

Qy 1115 TTACTGATCCCTTGAGCTGAGTACTTCTTGAGAGAGAAATTAAGATATAGCATCT 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 725 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 666

Qy 1175 CCGCCAAAGTCGAGCAAGGAATCAACGAACTAATGAATATCCATGGTGTGAGCT 1234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 606

Qy 1235 ACCCTCTCTTATTAACGATATCAACAGCTCTCAAGAGCTCAATAGCTCGGTGACT 1294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 ANNNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 546

Qy 1295 TGATTACCCCTTGATTATACGAAAGAACCTCTAAGATATCTACACAGCAATGAGA 1354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 486

Qy 1355 GAAAGAACTTTATCAACAAGATCAAGAGAGATCAAAATGAGAGAAAGAAATTGAGA 1414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 426

Qy 1415 GTGACAGAAAGATTACGAGACCGCAGCAAAAGTCTAAGCATCTCAAAAGATGTG 1474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 366

Qy 1475 AAAAGTCTGTAAGAGATCTATGATTCCAATTTCAACATTAACATGACCTGACCAACT 1534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 NNNNNANNNANNNANNNANNNANNNANNNANNNANNNANNNNNNNNNNNNN 306

Qy 1535 TCGAGAAATGATGGGAAAGCGTACTCTTACAAAGTGAGAGAACTGCACACCTATA 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 246

Qy 1595 CCTTGCATCTATGAGATTTCTAAGCATTAATCTTGAGAGCTCACCAAGCTCTTAAGT 1654

```

```

Db 245 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 186

Qy 1655 ATATGAGAGACTATCTCTGCGGAACATGTTGTGAGAAAGACTAAAGATTACAAGA 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126

Qy 1715 ATCTCATAGTAAGATGAGAAACGAGATCGAGAGCC 1750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 90

```

```

RESULT 40
BM656118      503 bp      mRNA      linear      EST 26-FEB-2002
LOCUS
DEFINITION 17000687387385 A.Gam.ad.cdNAL Anopheles gambiae cdNA clone
VERSION 19600449658190 5', mRNA sequence.
ACCESSION BM656118
KEYWORDS BM656118.1 GI:18955629
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS Holt R.A., Jin J.-J., Murphy S.D., Evans C.A., Kraft C.L., Charlab
R., Collins F.H., Venter J.C., and Hoffman S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltR@celera.com
Plate: ND01004HNK row: K column: 20
Seq primer: M13 Reverse.

```

```

FEATURES
source
location/qualifiers
1..503
/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449658190"
/clone_lib="A.Gam.ad.cdNAL"
/dev_stage="Adult"
/lab_host="DH10b"
/note="Vector: pSport1; site_1: SalI; site_2: NotI; whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."
```

```

BASE COUNT      300 a      124 c      72 g      7 t
ORIGIN
Query Match      1.6%: Score 77.2; DB 13; Length 503;
Best Local Similarity 51.0%; Pred. No. 9.9e-07;
Matches 209; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

```

```

Qy 693 CGAACATCAAGACATGTTGGAAAGATGAAATATTATTTAAAGATAAGAACAC 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 CACACACACACACACACACACACACACACACACACACACACACACACACACAG 148

Qy 753 CATCGAGAACATTAAAGCTGATCGAAGATCCAAAAGACCATAGCAAAAATAAGAA 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 CACACACACACACACACACACACACACACACACACACACACACACACACACAA 208

Qy 813 TGCAACCAAGGAGAGAAAGAGAGTGTACAGGCCCAAGTACGACCTGTCCATTA 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 CAGAACCAACAAACAGAACAGAACCAAGAACCAAGAACCAAGAACCAAGAACCA 268

```


Qy 873 TACAAACGCTTGAGAGCCCATACCTCATACGCTAGTGAGAGCGCATAGACAC 932
 Db 269 GAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 328
 Qy 933 CCTCAAGAAAGATGAATATATCAAGAACTGCTCGAAGATTATGAATTAAGATCC 992
 Db 329 GAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 385
 Qy 993 TCCGACGACCACTCTGGAGACACCCCTACACGCTGCTGGAGACAGAACAGATAGA 1052
 Db 386 CAAGAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 445
 Qy 1053 GGAGCAGCAGAACAGATCAAGAGATCGCCCAAAACCATTAAGTCAACA 1102
 Db 446 CAAGAACAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 495

RESULT 41
 BM159366 344 bp mRNA linear EST 04-DEC-2001
 LOCUS EST561889 PyBS Plasmodium yoelii yoelii cDNA clone PYCJD17 5' end,
 DEFINITION mRNA sequence.
 BM159366
 ACCESSION BM159366.1 GI:17305047
 VERSION EST.
 KEYWORDS Plasmodium yoelii yoelii.
 SOURCE Plasmodium yoelii yoelii.
 ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 344)
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B.,
 Fraser, C.M., and Carucci, D.D.
 JOURNAL Plasmodium yoelii EST project at TIGR
 COMMENT Unpublished (2001)
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-530-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.
 Location/Qualifiers
 1..344
 /organism="Plasmodium yoelii yoelii"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJD17"
 /clone_1lb="PyBS"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with PY17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidinium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase, and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZap arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HybridZap vector and plasmid DNA
 isolated."

BASE COUNT
 ORIGIN

145 a 65 c 59 g 75 t

Query Match 1.5%; Score 74.4; DB 13; Length 344;
 Best Local Similarity 57.9%; Pred. No. 3.9e-06;
 Matches 132; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Qy 1861 CAAGTCAGAGAGGTGCTCCTCATGAGCAAGATTGATGAACTCAAGAACTCAATTCATT 1920
 Db 2 CAATTCACAAAATCTTATTATTAACAACAAATTTGACCAATTAATTAACCTGAAGATCT 61
 Qy 1921 CTGAACAGCTGAGGATTAACATTAATATATACATGCCCAATAGTATTAAGCAGGAAT 1980
 Db 62 TTAAACAAGCTCAATTAAGCAACATATATGTTCCAAAAACATACGATATGAAGGA 121
 Qy 1981 AAGCAGAACCATATATACCTCATGCTACTCAGAGAGATAGACAAACTGAAGATGTTCC 2040
 Db 122 AAACCGAAGACATATATTTAATAGCTGTAAAGAAAGATTGACAGACTGTGCCAATTT 181
 Qy 2041 ATGCCCAAGTCAGAGCCGTGATCAACGAAGAACAGAACAGACTTAA 2088
 Db 182 ATTCCAAAATCGAAGATGATTGCTTAAAGAGAGAAAGAAATGGA 229

RESULT 42
 A2813205 540 bp DNA linear GSS 20-FEB-2001
 LOCUS ZM0080C24F Mouse 10kb plasmid UGCC1M library Mus musculus genomic
 DEFINITION clone UGCC2M0080C24 F, DNA sequence.
 A2813205
 ACCESSION A2813205.1 GI:12983113
 VERSION GSS.
 KEYWORDS house mouse.
 SOURCE Mus musculus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 540)
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,
 Islam, H., Longacre, S., Mahmoud, M., Meene, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,
 and Wright, D., Weiss, R.
 JOURNAL Mouse whole genome scaffolding with paired end reads from 10kb
 COMMENT Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0080 row: C column: 24
 Seq primer: CGTTGTAAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 540.
 Location/Qualifiers
 1..540
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGCC2M0080C24"
 /clone_1lb="Mouse 10kb plasmid UGCC1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to

FEATURES
 source

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g1473214|g1473214|p129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 284 a 54 c 151 g 51 t
ORIGIN

Query Match 1.5%; Score 73.8; DB 17; Length 540;
Best Local Similarity 49.9%; Pred. No. 6e-06;
Matches 186; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

Qy 731 ATATTAAAAAGATTAAGAGACCATCGAGACATTACGCTGATCGAAGATCCAAA 790
Db 115 AGATGAG 174
Qy 791 AGACCATGACAAAATTAAGATGCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 850
Db 175 AGAAG 234
Qy 851 CCCAGTACGACCTGTCTATCTTAACAAACAGCTGAAGAGAGAGAGAGAGAGAG 910
Db 235 AGAAG 294
Qy 911 TACTGAG 970
Db 295 AGAAG 354
Qy 971 AGATTAAAGAAATTAAGATCTCTCGCCAGCACTCTGAGAGAGAGAGAGAGAG 1030
Db 355 AGAAG 414
Qy 1031 TGACAG 1090
Db 415 AGAAG 474
Qy 1091 TTAAGTCAACAT 1103
Db 475 GCCAGAGCAATAT 487

RESULT 43
BM415494/c 948 bp mRNA linear EST 28-JAN-2002

LOCUS OP20570 Mixed Stage EST's from Globodera pallida, the potato cyst nematode Globodera pallida cDNA, mRNA sequence.

ACCESSION BM415494.1 GI:18382210

VERSION 1
KEYWORDS EST.

SOURCE Globodera pallida.

ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchima;

REFERENCE 1 (bases 1 to 948)
Heer, J., Sosinski, B., Pokrzywa, R. M., Warty, A. and Opperman, C.

AUTHORS Mixed Stage EST's from Globodera pallida, the potato cyst nematode unpublished (2001)

JOURNAL COMMENT: Opperman, C

Center for the Biology of Nematode Parasitism
NC State University, JACR-Rothamsted
Campus Box 7616; Raleigh, NC 27695, USA

TELE 919.515.6699
FAX 919.515.9500

EMAIL: warthog@unity.ncsu.edu
GTL1-4PCN.F.H07.PCN.4.F.060.ab1.

FEATURES
Location/Qualifiers
1..948

source
/organism="Globodera pallida"
/db_xref="taxon:36090"

/clone_11b="Mixed Stage EST's from Globodera pallida, the potato cyst nematode"
/note="Vector: lambda GT11. This is a collaborative effort between JACR-Rothamsted and North Carolina State University. The library was constructed from mixed stage G. pallida in lambda GT11 by Paul Burroughs, JACR-Rothamsted."

BASE COUNT 16 a 89 c 20 g 74 t 79 others
ORIGIN

Query Match 1.5%; Score 73.4; DB 13; Length 948;
Best Local Similarity 40.3%; Pred. No. 8.4e-06;
Matches 363; Conservative 0; Mismatches 532; Indels 6; Gaps 1;

Qy 694 GCAACATCAAGCAGCAATGTGGAAAGATGCAAGTTATTAAGAAAGAAAGACC 753
Db 920 GAAAG 861
Qy 754 ATCGAGAACATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 813
Db 860 AGAAG 801
Qy 814 GCAACCAAG 873
Db 800 ACAACCAAAANNN 741
Qy 874 AACCAACAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 933
Db 740 AAG 681
Qy 934 CTCAG 993
Db 680 AAAAAAG 621
Qy 994 CCGCCAGCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1053
Db 620 AAAAAAG 561
Qy 1054 GAGCAG 1113
Db 560 AAGGAG 501
Qy 1114 TTTACTGATCCCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1173
Db 500 AAAAAAG 447
Qy 1174 TCCGCCAAG 1233
Db 446 AAAAAAG 387
Qy 1234 TACCTCTGTCTTAAAGAGATCAACAGCTCTCAACGAGCTCAATAGCTTGGTG 1293
Db 386 AAAAAAG 327
Qy 1294 TTGATTAAACCCCTTGATTAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1353
Db 326 NNNNTTTTTCCTTCCCTTCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 267
Qy 1354 AGAAG 1413
Db 266 AAAAAAG 207
Qy 1414 AGTGACAG 1473
Db 206 AAAAAAG 147
Qy 1474 GAAAG 1533
Db 146 TACAG 87
Qy 1534 TTGAG 1593
Db 86 AAAAAAG 27

OY 1594 A 1594
Db 26 A 26

RESULT 44
BE420745
LOCUS BE420745 1885 bp mRNA linear EST 24-JUL-2000
DEFINITION HMM002.B02 ITRC HMM Barley Leaf Library Hordum vulgare cDNA clone
HMM002.B02, mRNA sequence.
ACCESSION BE420745.1 GI:9418588
VERSION BE420745.1
KEYWORDS EST.
SOURCE Hordum vulgare.
ORGANISM Hordum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 1885)
REFERENCE
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Landeridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
Pechioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sorrells,M., Warburton,M., and Wenzel,G.
International Triticeae EST Cooperative (ITRC): Production of
Expressed Sequence Tags for Species of the Triticeae
Unpublished (2000)
JOURNAL
COMMENT Contact: Hermann RG
Botanisches Institut der LMU
Menzinger Str. 67, D-80638 Muenchen GERMANY
Fax: 49 30 171683
Email: hermann@botanik.biologie.uni-muenchen.de
International Triticeae EST Cooperative (ITRC)
http://wheat.pw.usda.gov/genome.
LOCATION/Qualifiers
FEATURES
source 1.1885
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HMM002.B02"
/clone_lib="ITRC HMM Barley Leaf Library"
/tissue_type="leaf"
/dev_stage="14 day old"
/note="Vector: pBluescriptSk(-); 850 bp average insert
size."
BASE COUNT 1138 a 219 c 212 g 176 t 140 others
ORIGIN

Query Match 1.4%; Score 71.6; DB 10; Length 1885;
Best Local Similarity 37.6%; Pred. No. 2.6e-05;
Matches 362; Conservative 0; Mismatches 602; Indels 0; Gaps 0;

OY 631 AATATCAGACCCACGAGTGTGGACGATTTGAAGAAGTGTCTTCGATATCGCAGCCT 690
Db 892 AA 951
OY 691 CTCGACACATCAAGCAGCATGTGGAAAGATGAGATTATTTAAAGAATTAAGAG 750
Db 952 NNN 1011
OY 751 ACCATGAGAACATTAAGAGTGTGAGAAATCCAAAGAACATAGACAAATAAG 810
Db 1012 AA 1071
OY 811 AATGCAACACAGAGAGAGAAAGAAAGAGTTGTACAGGCCCAAGTACGACTGTCCATC 870
Db 1072 AA 1131
OY 871 TATAACAACAGCTTGAGAGGCCATTAACCTCATCGCTACTCGAGAGCCCATGAC 930
Db 1132 AA 1191

OY 931 ACCCTCAGAGAAGTGAATAATATCAAGAAGCTCGACAGATTATGAATTAGAAAT 990
Db 1192 AA 1251
OY 991 CCTCCGCGACCAACTCTGGGAACACCCCTTACACGCTGCTGGACAGAAAGATG 1050
Db 1252 AA 1311
OY 1051 GAGAGCAGCAGAGAAGAGATCAAGAGATCGCAAAACCTTAAGTTCAATAGATTCT 1110
Db 1312 AA 1371
OY 1111 CTCTTACTGATCCCTTGTGGCTGAGTACTTGTGAGAGAGAATTAAGATTAGAC 1170
Db 1372 AA 1431
OY 1171 ATCTCCGCCAAAGTCGAGCAAGAAATCAACCGAACCTTAATATCCCATTTGGTGTG 1230
Db 1432 AA 1491
OY 1231 ACCTACCCCTGTCTTATACGATATCAACAGCGCTCTCAAGAGCTCAATAGCTTGGT 1290
Db 1492 AA 1351
OY 1291 GACTTGATTACCCCTTGTGTTATACGAAAGAACCTTGAATATCTACACAGCAAT 1350
Db 1552 AA 1611
OY 1351 GAGAGAAAGAGATTATCAACGAAATCAAGAGAGATCAAAATTTGAGAAAGAAAT 1410
Db 1612 AA 1671
OY 1411 GAGAGTGACAAAGAAAGTTTCGAGACCGCAGCAAAAGCTTAACGATCTACTAAGAG 1470
Db 1672 AA 1731
OY 1471 TATGAAAAGCTGCTGAACGAGATCTATGTTCCAAATTCACANTTAACGACGCTGACC 1530
Db 1732 NAAA 1791
OY 1531 AACCTGAGAAATGATGGGAAACGCTACTTACAAAGTGAGAAATGACACACAT 1590
Db 1792 AA 1851
OY 1591 AATA 1594
Db 1852 AAAA 1855

RESULT 45
BH040136/c 691 bp DNA linear GSS 17-JUL-2001
LOCUS RPCI-24-255A23.TV RPCI-24 Mus musculus genomic clone RPCI-24-255A23
DEFINITION , DNA sequence.
ACCESSION BH040136
VERSION BH040136.1 GI:14818806
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsedaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P., and Fraser,C.M.
Titled
JOURNAL Mouse BAC End Sequences from Library RPCI-24
COMMENT Unpublished (1999)
Other GSSs: RPCI-24-255A23.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA


```

RESULT 47
A2548467 908 bp DNA linear GSS 14-NOV-2000
LOCUS ENTEK30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic DNA sequence.
ACCESSION A2548467
VERSION A2548467.1 GI:11172102
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 908)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HML:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 17
High quality sequence stop: 828.
FEATURES
source
1.908
Location/Qualifiers
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/Note="Vector: pHOSt1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."
```

```

BASE COUNT 434 a 16 c 289 g 169 t
ORIGIN
```

```

Query Match 1.4%; Score 68.6; DB 17; Length 908;
Best Local Similarity 45.1%; Pred. No. 0.0001;
Matches 343; Conservative 0; Mismatches 409; Indels 9; Gaps 2;
```

```

QY 1338 CTACACAGACATGAGAGAAAGATTTCACGAATCAAGAGAGATCAAAATGGA 1397
    || || || || || || || || || || || || || || || || || || ||
Db 7 CTGGAAGAGAGAGAGAGAGATGATGAAGATGATGATGAAGAGATGATGAAG 66
QY 1398 GAAGAAGAAATAGAGAGTACAGAAAGTTACGAAGCCGACGAAAGTCAACGA 1457
    || || || || || || || || || || || || || || || || || || ||
Db 67 GAAGATGATGATGATGAAGAGATGATGATGAAGAGATGATGAAGAGATGATGA 126
QY 1458 TATCACTAAAGAGTATGAAGAGCTGGAACGAGATCTATGATTCGAATTCACAAATTA 1517
    || || || || || || || || || || || || || || || || || || ||
Db 127 GAAGAAGATGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAG 186
QY 1518 CATGACCTGACCACTTGCAGAAATGATGGAAAACGCT---ACTCTTCAAAAGTGA 1574
    || || || || || || || || || || || || || || || || || || ||
Db 187 GATGATGATGATGAAGAGAGATGATGAAGATGATGAAGATGATGAAGAGATG 246
QY 1575 GAACGTGACACACCAATATACCTTTCATCTGAGATTTCAACATATATTTGAGAA 1634
    || || || || || || || || || || || || || || || || || || ||
```

```

Db 247 GAATATGAATTAAGAGATGATGATGATGAAGAGAGATGATGATGAAGAGAA 306
QY 1635 GCTCACCAAGACTCTTATGATATATGAGAGACTATTTCTCGGACACTTGTGGAGAA 1694
    || || || || || || || || || || || || || || || || || || ||
Db 307 GATGATGATGATGAAGATGATGATGAAGAGAGATGATGAAGAGATGATGAAG 366
QY 1695 AGAAGTAAAGTATTAAGAGATTCATTAAGATCGAAAACGAGATCGAGACG----- 1749
    || || || || || || || || || || || || || || || || || || ||
Db 367 GATGATGATGATGAAGAGAGATGATGATGAAGATGATGAAGAGATGATGAAG 426
QY 1750 -CTTGTGAGACATTAAGAGAGATGAGAGAGAGTGTGAGAGAGATGATGAAGAG 1808
    || || || || || || || || || || || || || || || || || || ||
Db 427 TATGATTAAGAGATGATGATGAAGAGAGATGATGATGAAGAGAGATGATGAAG 486
QY 1809 CGAAAATTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1868
    || || || || || || || || || || || || || || || || || || ||
Db 487 GATGAAGATGATGATGAAGAGAGATGATGATGATGATGATGATGATGATGATG 546
QY 1869 GAAGGTGCTCTGATGAGAGATGATGATGATGATGATGATGATGATGATGATG 1928
    || || || || || || || || || || || || || || || || || || ||
Db 547 GATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 606
QY 1929 CGTGAAGTTAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1988
    || || || || || || || || || || || || || || || || || || ||
Db 607 GATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 666
QY 1989 ACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2048
    || || || || || || || || || || || || || || || || || || ||
Db 667 GATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 726
QY 2049 AGTCGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2089
    || || || || || || || || || || || || || || || || || || ||
Db 727 GATGAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 767
```

```

RESULT 48
Bj420171 544 bp mRNA linear EST 10-MAR-2002
LOCUS Bj420171 Dictyostelium discoideum cDNA library, VF Dictyostelium
DEFINITION dictyostelium cDNA clone ddv38p08 5', mRNA sequence.
ACCESSION Bj420171
VERSION Bj420171.1 GI:19332411
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 544)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative
stage
JOURNAL Unpublished (2002)
CONTACT: Tadasi Shin-i
CENTER For Genetic Resource Information
NATIONAL Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
TEL: 81-559-81-6856
FAX: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
source
1.544
Location/Qualifiers
/organism="Dictyostelium discoideum"
/db_xref="taxon:44689"
/clone_lib="ddv38p08"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
/dev-stage="growth phase"
```

```

BASE COUNT 364 a 39 c 94 g 47 t
ORIGIN
Query Match 1.4%; Score 68.4; DB 13; Length 544;
Best Local Similarity 47.9%; Pred. No. 0.0001;
Matches 229; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
```


TITLE	Human gene number estimate provided by genome wide analysis using
JOURNAL	Tetraodon nigroviridis DNA sequence
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 907) Roest-Crollius H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 907)
AUTHORS	Genoscope.
TITLE	Direct Submision
JOURNAL	Submitted (12-APR-2000)
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .
DATE	20000412

BASE COUNT	27 a	30 c	22 g	759 t	69 others
ORIGIN					

Query Match	1.4%	Score 68	DB 17	Length 907
Best Local Similarity	39.2%	Pred. No. 0.00014		
Matches 337; Conservative	41;	Mismatches 482;	Indels 0;	Gaps 0

[illegible]

OY	1285	TGATTAAACCCCTGCATTATACGAAGAACCCTTAAGAAATTCATACACAGACAAATGAGA	1354
Db	268	AA	209
OY	1355	GAAGACACTTATATACACCAATCAAGAGAGATCAAAATTGAGAGAAGAAAATTGAGA	1414
Db	208	AA	149
OY	1415	GTGACACAAAGTAGTACGAGACCGCAGCAAAAGTCTAAAGCATTCCTAAAGAGTATG	1474
Db	148	AA	89
OY	1475	AAAAGCTGCTGACGAGATCTATGATTCCAAATTCACATATACATGCACCTGACCAACT	1534
Db	88	AA	29
OY	1535	TCGAGAAAATGATGGGAAAA	1554
Db	28	AAAAAAAAAAAAAAAAAAAAA	9

RESULT 51			
LOCUS	Bj360881		
DEFINITION	Bj360881 Dictyostelium discoideum cDNA library, CF Dictyostelium discoideum cDNA clone ddc8k12.5', mRNA sequence.	494 bp	mRNA, linear EST 07-MAR-2002

ACCESSION	BJ360881
VERSION	BJ360881.1
	GI:19260476

KEYWORDS	EST.
SOURCE	Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum
Eukaryota: Mycetozoa: Dictyostelida: Dictyostelium

REFERENCE

1 (bases 1 to 494)

AUTHORS Urushidaira, H., Tanaka, I., Konishi, I. and Saito, T., I.

TITLE Full length cDNA of Dictyostelium discoideum at the culmination

stage
Unpublished (2002)
JOURNAL

COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information

National Institute of Genetics
1111 West Michigan
Chicago 411-8510 T2000

1111 Iala, Mishima, SHIZUOKA 411-0240, Japan
Tel: 81-559-81-6856

Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES	Location/Qualifiers
source	1..494

```
/organism="Dictyostelium discoideum"
/strain="AY4"
```

```
/db_xref="taxon:44689"
```

```
/clone_1lib="Dictyostellum discoideum cDNA library, CF"
```

```
/sex="mat A"  
/dev stage="Culmination"
```

BASE COUNT	279 a	129 c	34 g	48 t	4 others
------------	-------	-------	------	------	----------

UNION

Query Match	1.48; Score 67.2; DB 13; Length 494;
Best Local Similarity	50.58; Pred. No. 0.00019;

Matches	162;	Conservative	0;	Mismatches	159;	Indels	0;	Gaps	
---------	------	--------------	----	------------	------	--------	----	------	--

QY 1223 ATGGTGTGACGTACCCCTCTGTCTTATAACGATATCACACACGCTCTCAACGAGCTCATATA 1282

Db 128 ATGTTGTAACGAACAATTCAACAACAACAAGAACAACAACAAGTACACAACAACCT 187

QY 1283 GCTTCGGTACCTTGATTAAACCCCTTCGATTATACGAAGAACCCTTAAGAATATCTACA 1342

[illegible][illegible]

1343 CAGACAAIGAGAGAAAGAGGIIITACACGGAAATCAAGCGGAGAGATCAAAAATGAGCGGA 1402

Db 248 ACACACACACAGCAGCACAACACACACACAGCACAACACACAGCACCACACACA 307

QY 1403 AGAAATTGAGAGTGACAGAAGTTACGAAGACCGCAGCAAAAGTCTAAACGATATCA 1462

[illegible]

DB	279	TGTCATTTAATGATCTCCTCACTAAGAAAGAAATTTAATGAAAAAATTTATTACAGATA	338
OY	1347	CATGAGAGAAAGAAAGTTTATTCACGAATACAGAGAGATCAAAATTGACAGACAA	1406
DB	339	TAAAGAAAGAAAGAAATTTATTCATTAATTAACATTAAAAAAATTTGATTGAGAGAAAAAAC	398
OY	1407	AATTGAGGTGACCAAGAAAGTTACGAGACCGACGCAAAAGTCTAAACGATATTCACATA	1466
DB	399	ATTAAATCACACAAAGAAACACATATAAAATAATTTACTTGAAGATTATGAA- AAGTCMAAAA	457
OY	1467	AGAGTATGAAAAAGCTGCAACGAGATCTGATTCCTCAAT	1508
DB	458	GGATTATGAGAGATTACTTGAAAAATTTATGAAATGAAAT	499
RESULT 53			
CNS02156			
LOCUS			
DEFINITION	Tetraodon nigroviridis genome survey sequence PUC-ori end of clone 22AF10 of library G from Tetraodon nigroviridis, genomic survey sequence.		
ACCESSION	AL176451		
VERSION	ALI76451.1		
KEYWORDS	GSS: genome survey sequence.		
SOURCE	Tetraodon nigroviridis.		
ORGANISM	Tetraodon nigroviridis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.		
REFERENCE	1 (bases 1 to 827) Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
AUTHORS	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
TITLE	Unpublished 2 (bases 1 to 827) Roest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
JOURNAL	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
REFERENCE	Unpublished 3 (bases 1 to 827) Roest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-APR-2000)		
COMMENT	This sequence is a single read and was generated as part of a large scale clone and sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .		
FEATURES	location/Qualifiers		
Source	1..827 /organism="Tetraodon nigroviridis" /db_xref="taxon:99883" /clone="22AF10" /clone_lib="G" /note="Genoscope sequence ID : C0AG224DC05SP1-end : PUC-ori"		
BASE COUNT	368 a 203 c 106 g 121 t 29 others		
ORIGIN			
Query Match	1.3%: Score 66.4; DB 17; Length 827;		
Best Local Similarity	43.2%: Pred. No. 0.00032;		
Matches	223; Conservative 16; Mismatches 277; Indels 0; Gaps 0;		
OY	1249	AACGATATCAACACGCTCTCAACGAGCTCATATAGCTTGCTGACTTGATTAACCCCTTC	1308
DB	209	ATTATCAACAATTAACACACACAAACAAATATTCACACATATACAGACGATTAACACAAAT	268
OY	1309	GATTATACGAAAGAACCTCTATAGATATTTACACAGACATGACAGAAAGATTATTC	1368

Db	1269	AACAAACAACACAAATTAACAACAACAATACAGSATAACACAATTAACAGCRTAAC	328
Oy	1369	AACGAATTCAGGACAGATCAAAATTTGAGAAGAAAGAAAATTTGAGGTGACAGAAGT	1428
Db	339	AACATTAACACACACACACAACTAATTAACAACAACAATTAACAGSATAACACAATTAACAGC	388
Oy	1429	TACGAAGACCCGAGCAAAAGTCTAANCGATATACCTAAAGACTATGAAGAAGCTGTGAAC	1488
Db	389	VATTAACAACAATTAACAACAATTAACAGTGTATTAACAACAATTAACACAACAATTAACAACAAT	448
Oy	1489	GAGATCTATGATTTCCAAATTCACAATAACATGACGTGACCAACTGTGAGAAAATGATG	1548
Db	449	RACACGATRACACRATGACAGCGATGRCACRACRACACAACAACAACAATTAAC	508
Oy	1549	GGAAGAAGCTACTCTTACAAAAGTGGAGAAAGCTGCACACACCAATTAATTCCTTTGCATCTAT	1608
Db	509	AGCGATTAACAGRATTAACAACAATTAACAGTGTATTAACAACAACAATTAACAGCGATTAACAACAT	568
Oy	1609	GAGAAATTCAGCAATTAATCTTGAGAAAGCTACACCAAGCTCTTAAGTATTTGAGAGCAAT	1668
Db	569	RACACATTAACACAGTGTATTAACAACAATTAACAGTGTATTAACACAAATTAACACAAACAAGTAA	628
Oy	1669	TCTCTGCGCAACATTTGTTGTGAGAGAAGAACTAAGATTTTACAGAATCTCATAGTAG	1728
Db	629	CAGCATTAACAACAATTAACAGCRATTAACAGCAATTAACAACAACAACAATCTCAGACGACG	688
Oy	1729	ATCGAAAACGAGATCGAAGCGTTGTTGGAACAATT	1764
Db	689	TTTAAAGAAAACGTCTCCCTCTCCCTCGGACGTGT	724

RESULT	54
B0931326/c	
LOCUS	B0931326/c
DEFINITION	B0931326 1168 bp mRNA linear EST 21-AUG-2007 AGNCOURT_8929072 NIH_MGC_40 Homo sapiens CDNA clone IMAGE:5484392
ACCESSION	5' , mRNA sequence.
VERSION	B0931326
KEYWORDS	B0931326.1 GI:22346357
SOURCE	EST.
ORGANISM	human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 1168)
TITLE	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cga@bsr.femail.nih.gov
COMMENT	Tissue Procurement: DCM/DMP CDNA Library Preparation: Rudin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing Arr: Agencourt Bioscience Corporation Clone distribution: MCC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LINC2670 row= 1 column= 01 High quality sequence stop:237.

FEATURES
SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6484352"
/clone_id="N1H_MGC_40"
/clone_type="carcinoma, cell line"
/tissue_type="prostate, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pOT87; Site_1: XhoI; Site_2: EcoRI; CNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(5'). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).

```

BASE COUNT	70	a	77	c	74	g	503	t	444	others
ORIGIN										
Query Match	1.3%	Score 66;	DB 14;	Length 1168;						
Best Local Similarity	23.5%	Pred. No.	0.00043							
Matches 207;	Conservative	0;	Matches 673;	Indels	0;	Gaps	0;			

OY	717	AAGATGGAAGCTTTATTATTAATAAAGAATPAGAACCATCGAACAATTACAGACTCAT	776
Dd	1137	AA	1078
OY	777	GGAAGAAATCAAAAAGACCATTAGACAAAAATPAAGATGCACCAAGAGAGAAAAAGAA	836
Dd	1077	AA	1018
OY	837	GAACTTGTACCAAGGCCAGTACGACCTCCATCTATPAACAACAGCTTGAAAGAGCCCA	896
Dd	1017	NAA	958
OY	897	TAACTTCATCAGCTGACTGGAGAACGCATGACACCCCTCAAGAAGATGAAATATCAA	956
Dd	957	NN	898
OY	957	AGAACTGCTCGACAAGATTATGAATTPAAGATCTCCGCCAGCCAATCTGGGACAC	1016
Dd	897	AA	838
OY	1017	CCCTAACACGCTGCTGGACACAGAACAGATAGAGAGCACGGAAGAAGATCAAGA	1076
Dd	837	AA	778
OY	1077	GATCGCCAAAACCATTAAGTTCACATAGATTCTCTTACTGATCCCCCTTGACCTGA	1136
Dd	777	NNANNNAANNNNNNNNNNNNNNNNANAANNAANNNNNNNNNNNNNNNNNNANAN	718
OY	1137	GTACTACTTGAGAGAGAGAAATAGATATATGACATCTCCGCCAAAGTCGACACAAAGA	1196
Dd	717	NN	658
OY	1197	ATCAACGCACTTAATGATATATCCCAATGGTGTGACGTACCTCTGTCTATPAAGATAT	1256
Dd	657	NAA	598
OY	1257	CAACAACGCTCTCAACGAGCTCAATAGCTTGGTGACTTGATTAAACCCCTTGATTATAC	1316
Dd	597	NN	538
OY	1317	GAAAGAACCCCTCTAAGAATATCTACACAGACATAGAGAAAGAGTTATCAACGAAT	1376
Dd	537	NN	478
OY	1377	CAAGGAGAAAGATCAAAATTGAGAGAGAGAAATTGAGAGTGACAGAAAGATTACGAGA	1436
Dd	477	NN	418
OY	1437	CCGACGCAAAAGTCAAAGATATCACTAAAGATGATAAAAGTCTGGAACGAGATCTA	1496
Dd	417	AA	358
OY	1497	TGATTCCAAATTCACAATTAACATGACCTGACCACTTCGAGAAATGATGGGAAAAAG	1556
Dd	357	NN	298
OY	1557	GTACTTACAAAGTGAGAGAACTGACACACCATATATAC	1596
Dd	297	NNNAAAGNNNGATGAAGAAAGACACACCCAGAGACACAGC	258

urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1020 Col=16 Row=G, DNA sequence.

ACCESSION

AZ191902

VERSION

AZ191902.1 GI:8375081

KEYWORDS

GSS.

SOURCE

Strongylocentrotus purpuratus.

ORGANIS

Strongylocentrotus purpuratus.

Strongylocentrotus purpuratus

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euarchinoidea; Echinacea; Echinoidea;

Echinoidea; Euechinoidea; Echinacea; Echinoida;
Echinoflorales; Echinoflorales; Echinoflorales

תוספת

Strongylocentrotidae; Strongylocentrotus.

D _b	282	AAAAAAAAATTAATTATTCACAAWAWTATAATAGTAAGTATGTAAGTGATATCTGCWCGGTG	223
Q _y	1753	GTTGAGAACCATTGAAGAGGTGAAGAACGTTGTTGGACAACAAGCTTRCAGAAAAGCGAA	1812
D _b	222	TATATAAAAATATATTAATTAATTATATRAAAMATATAATAAWTAAAAAARFTGTWMAAGRAGAT	163
Q _y	1813	AATAAACCCAGATGAGAAGA	1831
D _b	162	AWTAAWNTTAAAAAATAAAAAA	144

Search completed: March 31, 2003, 04:25:07
Job time : 4964 secs